(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
15 March 2001 (15.03.2001)

PCT

(10) International Publication Number
WO 01/18250 A2

(51) International Patent Classification[7]: C12Q 1/68, C07K 14/47

(21) International Application Number: PCT/US00/24503

(22) International Filing Date:
7 September 2000 (07.09.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/153,357	10 September 1999 (10.09.1999)	US
60/220,947	26 July 2000 (26.07.2000)	US
60/225,724	16 August 2000 (16.08.2000)	US

(71) Applicants: **WHITEHEAD INSTITUTE FOR BIOMEDICAL RESEARCH** [US/US]; Nine Cambridge Center, Cambridge, MA 02142 (US). **MILLENNIUM PHARMACEUTICALS, INC.** [US/US]; 75 Sidney Street, Cambridge, MA 02139 (US).

(72) Inventors: **LANDER, Eric, S.**; 151 Bishop Allen Drive, Cambridge, MA 02139 (US). **GARGILL, Michele**; 1 Warwick Park #3, Cambridge, MA 02140 (US). **IRELAND, James, S.**; 1 Warwick Park #3, Cambridge, MA 02140 (US). **BOLK, Stacey**; 202 Baker Street #1, West Roxbury, MA 02132 (US). **DALEY, George, Q.**; 50 Young Road, Weston, MA 02193 (US). **MCCARTHY, Jeanette, J.**; 3625 Dupont Street, San Diego, CA 92106 (US).

(74) Agents: **HOGLE, Doreen, M.** et al.; Hamilton, Brook, Smith & Reynolds, P.C., Two Militia Drive, Lexington, MA 02421 (US).

(81) Designated States *(national)*: AU, CA, JP, MX.

(84) Designated States *(regional)*: European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published:
— *Without international search report and to be republished upon receipt of that report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/18250 A2

(54) Title: SINGLE NUCLEOTIDE POLYMORPHISMS IN GENES

(57) Abstract: The invention provides nucleic acid segments of the human genome, particularly nucleic acid segments from a gene, including polymorphic sites. Allele-specific primers and probes hybridizing to regions flanking or containing these sites are also provided. The nucleic acids, primers and probes are used in applications such as phenotype correlations, forensics, paternity testing, medicine and genetic analysis. A role for the thrombospondin gene(s) in vascular disease is also disclosed. Use of single nucleotide polymorphisms in the thrombospondin gene(s) for diagnosis, prediction of clinical course and treatment response, development of therapeutics and development of cell-culture-based and animal models for research and treatment are disclosed.

SINGLE NUCLEOTIDE POLYMORPHISMS IN GENES

BACKGROUND OF THE INVENTION

The genomes of all organisms undergo spontaneous mutation in the course of their continuing evolution, generating variant forms of progenitor nucleic acid sequences (Gusella, *Ann. Rev. Biochem. 55,* 831-854 (1986)). The variant form may confer an evolutionary advantage or disadvantage relative to a progenitor form, or may be neutral. In some instances, a variant form confers a lethal disadvantage and is not transmitted to subsequent generations of the organism. In other instances, a variant form confers an evolutionary advantage to the species and is eventually incorporated into the DNA of many or most members of the species and effectively becomes the progenitor form. In many instances, both progenitor and variant form(s) survive and co-exist in a species population. The coexistence of multiple forms of a sequence gives rise to polymorphisms.

Several different types of polymorphism have been reported. A restriction fragment length polymorphism (RFLP) is a variation in DNA sequence that alters the length of a restriction fragment (Botstein *et al., Am. J. Hum. Genet.* 32, 314-331 (1980)). The restriction fragment length polymorphism may create or delete a restriction site, thus changing the length of the restriction fragment. RFLPs have been widely used in human and animal genetic analyses (see WO 90/13668; W090/11369; Donis-Keller, *Cell 51,* 319-337 (1987); Lander *et al., Genetics 121,* 85-99 (1989)). When a heritable trait can be linked to a particular RFLP, the presence of the RFLP in an individual can be used to predict the likelihood that the animal will also exhibit the trait.

Other polymorphisms take the form of short tandem repeats (STRs) that include tandem di-, tri- and tetra-nucleotide repeated motifs. These tandem repeats are also referred to as variable number tandem repeat (VNTR) polymorphisms.

VNTRs have been used in identity and paternity analysis (US 5,075,217; Armour *et al.*, *FEBS Lett. 307*, 113-115 (1992); Horn *et al.*, W0 91/14003; Jeffreys, EP 370,719), and in a large number of genetic mapping studies.

Other polymorphisms take the form of single nucleotide variations between individuals of the same species. Such polymorphisms are far more frequent than RFLPs, STRs and VNTRs. Some single nucleotide polymorphisms (SNP) occur in protein-coding nucleic acid sequences (coding sequence SNP (cSNP)), in which case, one of the polymorphic forms may give rise to the expression of a defective or otherwise variant protein and, potentially, a genetic disease. Examples of genes in which polymorphisms within coding sequences give rise to genetic disease include β-globin (sickle cell anemia), apoE4 (Alzheimer's Disease), Factor V Leiden (thrombosis), and CFTR (cystic fibrosis). cSNPs can alter the codon sequence of the gene and therefore specify an alternative amino acid. Such changes are called "missense" when another amino acid is substituted, and "nonsense" when the alternative codon specifies a stop signal in protein translation. When the cSNP does not alter the amino acid specified the cSNP is called "silent".

Other single nucleotide polymorphisms occur in noncoding regions. Some of these polymorphisms may also result in defective protein expression (e.g., as a result of defective splicing). Other single nucleotide polymorphisms have no phenotypic effects. Single nucleotide polymorphisms can be used in the same manner as RFLPs and VNTRs, but offer several advantages. Single nucleotide polymorphisms occur with greater frequency and are spaced more uniformly throughout the genome than other forms of polymorphism. The greater frequency and uniformity of single nucleotide polymorphisms means that there is a greater probability that such a polymorphism will be found in close proximity to a genetic locus of interest than would be the case for other polymorphisms. The different forms of characterized single nucleotide polymorphisms are often easier to distinguish than other types of polymorphism (e.g., by use of assays employing allele-specific hybridization probes or primers).

Only a small percentage of the total repository of polymorphisms in humans and other organisms has been identified. The limited number of polymorphisms identified to date is due to the large amount of work required for their detection by

conventional methods. For example, a conventional approach to identifying polymorphisms might be to sequence the same stretch of DNA in a population of individuals by dideoxy sequencing. In this type of approach, the amount of work increases in proportion to both the length of sequence and the number of individuals in a population and becomes impractical for large stretches of DNA or large numbers of persons.

SUMMARY OF THE INVENTION

Work described herein pertains to the identification of polymorphisms which can predispose individuals to disease, by resequencing large numbers of genes in a large number of individuals. Various genes from a number of individuals have been resequenced as described herein, and SNPs in these genes have been discovered (see the Table and Fig. 3). Some of these SNPs are cSNPs which specify a different amino acid sequence, some of the SNPs are silent cSNPs and some of these cSNPs specify a stop signal in protein translation. Some of the identified SNPs were located in non-coding regions.

The invention relates to a gene which comprises a single nucleotide polymorphism at a specific location. In a particular embodiment the invention relates to the variant allele of a gene having a single nucleotide polymorphism, which variant allele differs from a reference allele by one nucleotide at the site(s) identified in the Table and Fig. 3. Complements of these nucleic acid sequences are also included. The nucleic acid molecules can be DNA or RNA, and can be double- or single-stranded. Nucleic acid molecules can be, for example, 5-10, 5-15, 10-20, 5-25, 10-30, 10-50 or 10-100 bases long.

The invention further provides allele-specific oligonucleotides that hybridize to the reference or variant allele of a gene comprising a single nucleotide polymorphism or to the complement thereof. These oligonucleotides can be probes or primers.

The invention further provides a method of analyzing a nucleic acid from an individual. The method determines which base is present at any one of the polymorphic sites shown in the Table and/or Fig. 3. Optionally, a set of bases occupying a set of the polymorphic sites shown in the Table and /or Fig. 3 is

determined. This type of analysis can be performed on a number of individuals, who are tested for the presence of a disease phenotype. The presence or absence of disease phenotype is then correlated with a base or set of bases present at the polymorphic site or sites in the individuals tested.

Thus, the invention further relates to a method of predicting the presence, absence, likelihood of the presence or absence, or severity of a particular phenotype or disorder associated with a particular genotype. The method comprises obtaining a nucleic acid sample from an individual and determining the identity of one or more bases (nucleotides) at polymorphic sites of genes described herein, wherein the presence of a particular base is correlated with a specified phenotype or disorder, thereby predicting the presence, absence, likelihood of the presence or absence, or severity of the phenotype or disorder in the individual.

The thrombospondins are a family of extracellular matrix (ECM) glycoproteins that modulate many cell behaviors including adhesion, migration, and proliferation. Thrombospondins (also known as thrombin sensitive proteins or TSPs) are large molecular weight glycoproteins composed of three identical disulfide-linked polypeptide chains. The results described herein also reveal an important association between alterations, particularly SNPs, in TSP genes, particularly TSP-1 and TSP-4, and vascular disease. In particular, SNPs in these genes which are associated with premature coronary artery disease (CAD)(or coronary heart disease) and myocardial infarction (MI) have been identified and represent a potentially vital marker of upstream biology influencing the complex process of atherosclerotic plaque generation and vulnerability.

Thus, the invention relates to the TSP gene SNPs identified as described herein, both singly and in combination, as well as to the use of these SNPs, and others in TSP genes, particularly those nearby in linkage disequilibrium with these SNPs, for diagnosis, prediction of clinical course and treatment response for vascular disease, development of new treatments for vascular disease based upon comparison of the variant and normal versions of the gene or gene product, and development of cell-culture based and animal models for research and treatment of vascular disease. The invention further relates to novel compounds and

pharmaceutical compositions for use in the diagnosis and treatment of such disorders. In preferred embodiments, the vascular disease is CAD or MI.

The invention relates to isolated nucleic acid molecules comprising all or a portion of the variant allele of TSP-1 (e.g., as exemplified by SEQ ID NO: 1), and to isolated nucleic acid molecules comprising all or a portion of the variant allele of TSP-4 (e.g., as exemplified by SEQ ID NO: 3). Preferred portions are at least 10 contiguous nucleotides and comprise the polymorphic site, e.g., a portion of SEQ ID NO: 1 which is at least 10 contiguous nucleotides and comprises the "G" at position 2210, or a portion of SEQ ID NO: 3 which is at least 10 contiguous nucleotides and comprises the "C" at position 1186. The invention further relates to isolated gene products, e.g., polypeptides or proteins, which are encoded by a nucleic acid molecule comprising all or a portion of the variant allele of TSP-1 or TSP-4 (e.g., SEQ ID NO: 1 or SEQ ID NO: 3, respectively). The invention also relates to nucleic acid molecules which hybridize to and/or share identity with the variant alleles identified herein (or their complements) and which also comprise the variant nucleotide at the SNP site.

The invention further relates to isolated proteins or polypeptides comprising all or a portion of the variant amino acid sequence of TSP-1 (e.g., as exemplified by SEQ ID NO: 2), and to isolated proteins or polypeptides comprising all or a portion of the variant amino acid sequence of TSP-4 (e.g., as exemplified by SEQ ID NO: 4). Preferred polypeptides are at least 10 contiguous amino acids and comprise the polymorphic amino acid, e.g., a portion of SEQ ID NO: 2 which is at least 10 contiguous amino acids and comprises the serine at residue 700, or a portion of SEQ ID NO: 4 which is at least 10 contiguous amino acids and comprises the proline at residue 387. The invention further relates to isolated nucleic acid molecules encoding such proteins and polypeptides, as well as to antibodies which bind, e.g., specifically, to such proteins and polypeptides.

The invention further relates to a method of diagnosing or aiding in the diagnosis of a disorder associated with the presence of one or more of (a) a G at nucleotide position 2210 of SEQ ID NO: 1; or (b) a C at nucleotide position 1186 of SEQ ID NO: 3 in an individual. The method comprises obtaining a nucleic acid sample from the individual and determining the nucleotide present at one or more of

the indicated nucleotide positions, wherein presence of one or more of (a) a G at nucleotide position 2210 of SEQ ID NO: 1; or (b) a C at nucleotide position 1186 of SEQ ID NO: 3 is indicative of increased likelihood of said disorder in the individual as compared with an appropriate control, *e.g.*, an individual having the reference nucleotide at one or more of said positions. In a particular embodiment the disorder is a vascular disease selected from the group consisting of atherosclerosis, coronary heart or artery disease, MI, stroke, peripheral vascular diseases, venous thromboembolism and pulmonary embolism. In a preferred embodiment, the vascular disease is selected from the group consisting of CAD and MI.

The invention further relates to a method of diagnosing or aiding in the diagnosis of a disorder associated with one or more of (a) a G at nucleotide position 2210 of SEQ ID NO: 1; or (b) a C at nucleotide position 1186 of SEQ ID NO: 3 in an individual. The method comprises obtaining a nucleic acid sample from the individual and determining the nucleotide present at one or more of the indicated nucleotide positions, wherein presence of one or more of (a) an A at nucleotide position 2210 of SEQ ID NO: 1; or (b) a G at nucleotide position 1186 of SEQ ID NO: 3 is indicative of decreased likelihood of said disorder in the individual as compared with an appropriate control, *e.g.*, an individual having the variant nucleotide at said position. In a particular embodiment the disorder is a vascular disease selected from the group consisting of atherosclerosis, coronary heart or artery disease, MI, stroke, peripheral vascular diseases, venous thromboembolism and pulmonary embolism. In a preferred embodiment, the vascular disease is selected from the group consisting of CAD and MI.

In one embodiment, the invention relates to a method for predicting the likelihood that an individual will have a vascular disease (or aiding in the diagnosis of a vascular disease), comprising the steps of obtaining a DNA sample from an individual to be assessed and determining the nucleotide present at one or more of nucleotide positions 2210 of SEQ ID NO: 1 or 1186 of SEQ ID NO: 3. The presence of the reference nucleotide at one or more of these positions indicates that the individual has a lower likelihood of having a vascular disease than an individual having the variant nucleotide at one or more of these positions, or a lower likelihood

of having severe symptomology. In a particular embodiment, the individual is an individual at risk for development of a vascular disease.

The invention further relates to a method of diagnosing or aiding in the diagnosis of a disorder associated with the presence of one or more of (a) a serine at amino acid position 700 of SEQ ID NO: 2; or (b) a proline at amino acid position 387 of SEQ ID NO: 4 in an individual. The method comprises obtaining a biological sample containing the TSP-1 and/or TSP-4 protein or relevant portion thereof from the individual and determining the amino acid present at one or more of the indicated amino acid positions, wherein presence of one or more of (a) a serine at amino acid position 700 of SEQ ID NO: 2; or (b) a proline at amino acid position 387 of SEQ ID NO: 4 is indicative of increased likelihood of said disorder in the individual as compared with an appropriate control, *e.g.*, an individual having the reference amino acid at one or more of said positions.

The invention further relates to a method of diagnosing or aiding in the diagnosis of a disorder associated with one or more of (a) a serine at amino acid position 700 of SEQ ID NO: 2; or (b) a proline at amino acid position 387 of SEQ ID NO: 4 in an individual. The method comprises obtaining a biological sample containing the TSP-1 and/or TSP-4 protein or relevant portion thereof from the individual and determining the amino acid present at one or more of the indicated amino acid positions, wherein presence of one or more of (a) an asparagine at amino acid position 700 of SEQ ID NO: 2; or (b) an alanine at amino acid position 387 of SEQ ID NO: 4 is indicative of decreased likelihood of said disorder in the individual as compared with an appropriate control, *e.g.*, an individual having the variant amino acid at one or more of said positions.

In one embodiment, the invention relates to a method for predicting the likelihood that an individual will have a vascular disease (or aiding in the diagnosis of a vascular disease), comprising the steps of obtaining a biological sample comprising the TSP-1 and/or TSP-4 protein or relevant portion thereof from an individual to be assessed and determining the amino acid present at one or more of amino acid positions 700 of SEQ ID NO: 2 or 387 of SEQ ID NO: 4. The presence of the reference amino acid at one or more of these positions indicates that the individual has a lower likelihood of having a vascular disease than an individual

having the variant amino acid at one or more of these positions, or a lower likelihood of having severe symptomology. In a particular embodiment, the individual is an individual at risk for development of a vascular disease.

In another embodiment, the invention relates to pharmaceutical compositions comprising a reference TSP-1 and/or TSP-4 gene or gene product, or active portion thereof, for use in the treatment of vascular diseases. The invention further relates to the use of agonists and antagonists of TSP-1 and TSP-4 activity for use in the treatment of vascular diseases. In a particular embodiment the vascular disease is selected from the group consisting of atherosclerosis, coronary heart or artery disease, MI, stroke, peripheral vascular diseases, venous thromboembolism and pulmonary embolism. In a preferred embodiment, the vascular disease is selected from the group consisting of CAD and MI.

BRIEF DESCRIPTION OF THE DRAWINGS

Figs. 1A-1D show the reference nucleotide (SEQ ID NO: 1) and amino acid (SEQ ID NO: 2) sequences for TSP-1.

Figs. 2A-2C show the reference nucleotide (SEQ ID NO: 3) and amino acid (SEQ ID NO: 4) sequences for TSP-4.

Fig. 3 shows a table providing detailed information about the SNPs identified herein. Column one shows the internal polymorphism identifier. Column two shows the accession number for the reference sequence in the TIGR database (http://www.tigr.org/tdb/hgi/searching/hgi_reports.html). Column three shows the nucleotide position for the SNP iste. Column four shows the gene in which the polymorphism was identified. Column five shows the polymorphic site and additional flanking sequence on each side of the polymorphism. Column six shows the type of mutation produced by the polymorphism. Columns seven and eight show the reference and alternate (variant) nucleotides, respectively, for the SNP. Columns nine and ten show the reference and alternate (variant) amino acids, respectively, encoded by the alleles of the gene.

DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to a gene which comprises a single nucleotide polymorphism (SNP) at a specific location. The gene which includes the SNP has at least two alleles, referred to herein as the reference allele and the variant allele. The reference allele (prototypical or wild type allele) has been designated arbitrarily and typically corresponds to the nucleotide sequence of the gene which has been deposited with GenBank or TIGR under a given Accession number. The variant allele differs from the reference allele by one nucleotide at the site(s) identified in the Table. The present invention also relates to variant alleles of the described genes and to complements of the variant alleles. The invention also relates to nucleic acid molecules which hybridize to and/or share identity with the variant alleles identified herein (or their complements) and which also comprise the variant nucleotide at the SNP site.

The invention further relates to portions of the variant alleles and portions of complements of the variant alleles which comprise (encompass) the site of the SNP and are at least 5 nucleotides in length. Portions can be, for example, 5-10, 5-15, 10-20, 5-25, 10-30, 10-50 or 10-100 bases long. For example, a portion of a variant allele which is 21 nucleotides in length includes the single nucleotide polymorphism (the nucleotide which differs from the reference allele at that site) and twenty additional nucleotides which flank the site in the variant allele. These nucleotides can be on one or both sides of the polymorphism. Polymorphisms which are the subject of this invention are defined in the Table with respect to the reference sequence deposited in GenBank or TIGR under the Accession number indicated. For example, the invention relates to a portion of a gene (e.g., AT3) having a nucleotide sequence as deposited in GenBank (e.g., U11270) comprising a single nucleotide polymorphism at a specific position (e.g., nucleotide 11918). The reference nucleotide for AT3 is shown in column 8, and the variant nucleotide is shown in column 9 of the Table. The nucleotide sequences of the invention can be double- or single-stranded.

The invention further provides allele-specific oligonucleotides that hybridize to the reference or variant allele of a gene comprising a single nucleotide

polymorphism or to the complement thereof. These oligonucleotides can be probes or primers.

The invention further provides a method of analyzing a nucleic acid from an individual. The method determines which base is present at any one of the polymorphic sites shown in the Table and/or Fig. 3. Optionally, a set of bases occupying a set of the polymorphic sites shown in the Table and/or Fig. 3 is determined. This type of analysis can be performed on a number of individuals, who are tested for the presence of a disease phenotype. The presence or absence of disease phenotype is then correlated with a base or set of bases present at the polymorphic site or sites in the individuals tested.

Thus, the invention further relates to a method of predicting the presence, absence, likelihood of the presence or absence, or severity of a particular phenotype or disorder associated with a particular genotype. The method comprises obtaining a nucleic acid sample from an individual and determining the identity of one or more bases (nucleotides) at polymorphic sites of genes described herein, wherein the presence of a particular base is correlated with a specified phenotype or disorder, thereby predicting the presence, absence, likelihood of the presence or absence, or severity of the phenotype or disorder in the individual.

DEFINITIONS

A nucleic acid molecule or oligonucleotide can be DNA or RNA, and single- or double-stranded. Nucleic acid molecules and oligonucleotides can be naturally occurring or synthetic, but are typically prepared by synthetic means. Preferred nucleic acid molecules and oligonucleotides of the invention include segments of DNA, or their complements, which include any one of the polymorphic sites shown in the Table. The segments can be between 5 and 250 bases, and, in specific embodiments, are between 5-10, 5-20, 10-20, 10-50, 20-50 or 10-100 bases. For example, the segment can be 21 bases. The polymorphic site can occur within any position of the segment. The segments can be from any of the allelic forms of DNA shown in the Table.

As used herein, the terms "nucleotide", "base" and "nucleic acid" are intended to be equivalent. The terms "nucleotide sequence", "nucleic acid sequence", "nucleic acid molecule" and "segment" are intended to be equivalent.

Hybridization probes are oligonucleotides which bind in a base-specific manner to a complementary strand of nucleic acid. Such probes include peptide nucleic acids, as described in Nielsen *et al.*, *Science* 254, 1497-1500 (1991). Probes can be any length suitable for specific hybridization to the target nucleic acid sequence. The most appropriate length of the probe may vary depending upon the hybridization method in which it is being used; for example, particular lengths may be more appropriate for use in microfabricated arrays, while other lengths may be more suitable for use in classical hybridization methods. Such optimizations are known to the skilled artisan. Suitable probes and primers can range from about 5 nucleotides to about 30 nucleotides in length. For example, probes and primers can be 5, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 25, 26, 28 or 30 nucleotides in length. The probe or primer preferably overlaps at least one polymorphic site occupied by any of the possible variant nucleotides. The nucleotide sequence can correspond to the coding sequence of the allele or to the complement of the coding sequence of the allele.

As used herein, the term "primer" refers to a single-stranded oligonucleotide which acts as a point of initiation of template-directed DNA synthesis under appropriate conditions (*e.g.*, in the presence of four different nucleoside triphosphates and an agent for polymerization, such as DNA or RNA polymerase or reverse transcriptase) in an appropriate buffer and at a suitable temperature. The appropriate length of a primer depends on the intended use of the primer, but typically ranges from 15 to 30 nucleotides. Short primer molecules generally require cooler temperatures to form sufficiently stable hybrid complexes with the template. A primer need not reflect the exact sequence of the template, but must be sufficiently complementary to hybridize with a template. The term primer site refers to the area of the target DNA to which a primer hybridizes. The term primer pair refers to a set of primers including a 5' (upstream) primer that hybridizes with the 5' end of the DNA sequence to be amplified and a 3' (downstream) primer that hybridizes with the complement of the 3' end of the sequence to be amplified.

As used herein, linkage describes the tendency of genes, alleles, loci or genetic markers to be inherited together as a result of their location on the same chromosome. It can be measured by percent recombination between the two genes, alleles, loci or genetic markers.

As used herein, polymorphism refers to the occurrence of two or more genetically determined alternative sequences or alleles in a population. A polymorphic marker or site is the locus at which divergence occurs. Preferred markers have at least two alleles, each occurring at frequency of greater than 1%, and more preferably greater than 10% or 20% of a selected population. A polymorphic locus may be as small as one base pair. Polymorphic markers include restriction fragment length polymorphisms, variable number of tandem repeats (VNTR's), hypervariable regions, minisatellites, dinucleotide repeats, trinucleotide repeats, tetranucleotide repeats, simple sequence repeats, and insertion elements such as Alu. The first identified allelic form is arbitrarily designated as the reference form and other allelic forms are designated as alternative or variant alleles. The allelic form occurring most frequently in a selected population is sometimes referred to as the wildtype form. Diploid organisms may be homozygous or heterozygous for allelic forms. A diallelic or biallelic polymorphism has two forms. A triallelic polymorphism has three forms.

Work described herein pertains to the resequencing of large numbers of genes in a large number of individuals to identify polymorphisms which can predispose individuals to disease. For example, polymorphisms in genes which are expressed in liver may predispose individuals to disorders of the liver. By altering amino acid sequence, SNPs may alter the function of the encoded proteins. The discovery of the SNP facilitates biochemical analysis of the variants and the development of assays to characterize the variants and to screen for pharmaceutical that would interact directly with one or another form of the protein. SNPs (including silent SNPs) also enable the development of specific DNA, RNA, or protein-based diagnostics that detect the presence or absence of the polymorphism in particular conditions.

A single nucleotide polymorphism occurs at a polymorphic site occupied by a single nucleotide, which is the site of variation between allelic sequences. The site

is usually preceded by and followed by highly conserved sequences of the allele (e.g., sequences that vary in less than 1/100 or 1/1000 members of the populations).

A single nucleotide polymorphism usually arises due to substitution of one nucleotide for another at the polymorphic site. A transition is the replacement of one purine by another purine or one pyrimidine by another pyrimidine. A transversion is the replacement of a purine by a pyrimidine or vice versa. Single nucleotide polymorphisms can also arise from a deletion of a nucleotide or an insertion of a nucleotide relative to a reference allele. Typically the polymorphic site is occupied by a base other than the reference base. For example, where the reference allele contains the base "T" at the polymorphic site, the altered allele can contain a "C", "G" or "A" at the polymorphic site.

The invention also relates to nucleic acid molecules which hybridize to the variant alleles identified herein (or their complements) and which also comprise the variant nucleotide at the SNP site. Hybridizations are usually performed under stringent conditions, for example, at a salt concentration of no more than 1 M and a temperature of at least 25°C. For example, conditions of 5X SSPE (750 mM NaCl, 50 mM NaPhosphate, 5 mM EDTA, pH 7.4) and a temperature of 25-30°C, or equivalent conditions, are suitable for allele-specific probe hybridizations. Equivalent conditions can be determined by varying one or more of the parameters given as an example, as known in the art, while maintaining a similar degree of identity or similarity between the target nucleotide sequence and the primer or probe used.

The invention also relates to nucleic acid molecules which share substantial sequence identity to the variant alleles identified herein (or their complements) and which also comprise the variant nucleotide at the SNP site. Particularly preferred are nucleic acid molecules and fragments which have at least about 60%, preferably at least about 70, 80 or 85%, more preferably at least about 90%, even more preferably at least about 95%, and most preferably at least about 98% identity with nucleic acid molecules described herein. The percent identity of two nucleotide or amino acid sequences can be determined by aligning the sequences for optimal comparison purposes (*e.g.*, gaps can be introduced in the sequence of a first sequence). The nucleotides or amino acids at corresponding positions are then

compared, and the percent identity between the two sequences is a function of the number of identical positions shared by the sequences (*i.e.*, % identity = # of identical positions/total # of positions x 100). In certain embodiments, the length of a sequence aligned for comparison purposes is at least 30%, preferably at least 40%, more preferably at least 60%, and even more preferably at least 70%, 80% or 90% of the length of the reference sequence. The actual comparison of the two sequences can be accomplished by well-known methods, for example, using a mathematical algorithm. A preferred, non-limiting example of such a mathematical algorithm is described in Karlin *et al.*, *Proc. Natl. Acad. Sci. USA, 90*:5873-5877 (1993). Such an algorithm is incorporated into the NBLAST and XBLAST programs (version 2.0) as described in Altschul *et al.*, *Nucleic Acids Res., 25*:389-3402 (1997). When utilizing BLAST and Gapped BLAST programs, the default parameters of the. respective programs (*e.g.*, NBLAST) can be used. See http://www.ncbi.nlm.nih.gov. In one embodiment, parameters for sequence comparison can be set at score=100, wordlength=12, or can be varied (*e.g.*, W=5 or W=20).

The term "isolated" is used herein to indicate that the material in question exists in a physical milieu distinct from that in which it occurs in nature. For example, an isolated nucleic acid of the invention may be substantially isolated with respect to the complex cellular milieu in which it naturally occurs. In some instances, the isolated material will form part of a composition (for example, a crude extract containing other substances), buffer system or reagent mix. In other circumstance, the material may be purified to essential homogeneity, for example as determined by PAGE or column chromatography such as HPLC. Preferably, an isolated nucleic acid comprises at least about 50, 80 or 90 percent (on a molar basis) of all macromolecular species present.

I. Novel Polymorphisms of the Invention

Some of the novel polymorphisms of the invention are shown in the Table. Columns one and two show designations for the indicated polymorphism. Column three shows the Genbank or TIGR Accession number for the wild type (or reference) allele. Column four shows the location of the polymorphic site in the nucleic acid

sequence with reference to the Genbank or TIGR sequence shown in column three. Column five shows common names for the gene in which the polymorphism is located. Column six shows the polymorphism and a portion of the 3' and 5' flanking sequence of the gene. Column seven shows the type of mutation; N, non-sense, S, silent, M, missense. Columns eight and nine show the reference and alternate nucleotides, respectively, at the polymorphic site. Columns ten and eleven show the reference and alternate amino acids, respectively, encoded by the reference and variant, respectively, alleles. Other novel polymorphisms of the invention are shown in Fig. 3.

II. Analysis of Polymorphisms

A. Preparation of Samples

Polymorphisms are detected in a target nucleic acid from an individual being analyzed. For assay of genomic DNA, virtually any biological sample (other than pure red blood cells) is suitable. For example, convenient tissue samples include whole blood, semen, saliva, tears, urine, fecal material, sweat, buccal, skin and hair. For assay of cDNA or mRNA, the tissue sample must be obtained from an organ in which the target nucleic acid is expressed. For example, if the target nucleic acid is a cytochrome P450, the liver is a suitable source.

Many of the methods described below require amplification of DNA from target samples. This can be accomplished by e.g., PCR. *See generally PCR Technology: Principles and Applications for DNA Amplification* (ed. H.A. Erlich, Freeman Press, NY, NY, 1992); *PCR Protocols: A Guide to Methods and Applications* (eds. Innis, *et al.*, Academic Press, San Diego, CA, 1990); Mattila *et al.*, *Nucleic Acids Res.* 19, 4967 (1991); Eckert *et al.*, *PCR Methods and Applications* 1, 17 (1991); *PCR* (eds. McPherson *et al.*, IRL Press, Oxford); and U.S. Patent 4,683,202.

Other suitable amplification methods include the ligase chain reaction (LCR) (see Wu and Wallace, *Genomics* 4, 560 (1989), Landegren *et al.*, *Science* 241, 1077 (1988), transcription amplification (Kwoh *et al.*, *Proc. Natl. Acad. Sci. USA* 86, 1173 (1989)), and self-sustained sequence replication (Guatelli *et al.*, *Proc. Nat. Acad. Sci. USA*, 87, 1874 (1990)) and nucleic acid based sequence amplification

(NASBA). The latter two amplification methods involve isothermal reactions based on isothermal transcription, which produce both single stranded RNA (ssRNA) and double stranded DNA (dsDNA) as the amplification products in a ratio of about 30 or 100 to 1, respectively.

B. Detection of Polymorphisms in Target DNA

There are two distinct types of analysis of target DNA for detecting polymorphisms. The first type of analysis, sometimes referred to as *de novo* characterization, is carried out to identify polymorphic sites not previously characterized (i.e., to identify new polymorphisms). This analysis compares target sequences in different individuals to identify points of variation, i.e., polymorphic sites. By analyzing groups of individuals representing the greatest ethnic diversity among humans and greatest breed and species variety in plants and animals, patterns characteristic of the most common alleles/haplotypes of the locus can be identified, and the frequencies of such alleles/haplotypes in the population can be determined. Additional allelic frequencies can be determined for subpopulations characterized by criteria such as geography, race, or gender. The *de novo* identification of polymorphisms of the invention is described in the Examples section. The second type of analysis determines which form(s) of a characterized (known) polymorphism are present in individuals under test. There are a variety of suitable procedures, which are discussed in turn.

1. Allele-Specific Probes

The design and use of allele-specific probes for analyzing polymorphisms is described by e.g., Saiki *et al.*, *Nature* 324, 163-166 (1986); Dattagupta, EP 235,726, Saiki, WO 89/11548. Allele-specific probes can be designed that hybridize to a segment of target DNA from one individual but do not hybridize to the corresponding segment from another individual due to the presence of different polymorphic forms in the respective segments from the two individuals. Hybridization conditions should be sufficiently stringent that there is a significant difference in hybridization intensity between alleles, and preferably an essentially binary response, whereby a probe hybridizes to only one of the alleles. Some probes

are designed to hybridize to a segment of target DNA such that the polymorphic site aligns with a central position (e.g., in a 15-mer at the 7 position; in a 16-mer, at either the 8 or 9 position) of the probe. This design of probe achieves good discrimination in hybridization between different allelic forms.

Allele-specific probes are often used in pairs, one member of a pair showing a perfect match to a reference form of a target sequence and the other member showing a perfect match to a variant form. Several pairs of probes can then be immobilized on the same support for simultaneous analysis of multiple polymorphisms within the same target sequence.

2. Tiling Arrays

The polymorphisms can also be identified by hybridization to nucleic acid arrays, some examples of which are described in WO 95/11995. One form of such arrays is described in the Examples section in connection with de novo identification of polymorphisms. The same array or a different array can be used for analysis of characterized polymorphisms. WO 95/11995 also describes subarrays that are optimized for detection of a variant form of a precharacterized polymorphism. Such a subarray contains probes designed to be complementary to a second reference sequence, which is an allelic variant of the first reference sequence. The second group of probes is designed by the same principles as described in the Examples, except that the probes exhibit complementarity to the second reference sequence. The inclusion of a second group (or further groups) can be particularly useful for analyzing short subsequences of the primary reference sequence in which multiple mutations are expected to occur within a short distance commensurate with the length of the probes (e.g., two or more mutations within 9 to 21 bases).

3. Allele-Specific Primers

An allele-specific primer hybridizes to a site on target DNA overlapping a polymorphism and only primes amplification of an allelic form to which the primer exhibits perfect complementarity. See Gibbs, *Nucleic Acid Res.* 17, 2427-2448 (1989). This primer is used in conjunction with a second primer which hybridizes at a distal site. Amplification proceeds from the two primers, resulting in a detectable

product which indicates the particular allelic form is present. A control is usually performed with a second pair of primers, one of which shows a single base mismatch at the polymorphic site and the other of which exhibits perfect complementarity to a distal site. The single-base mismatch prevents amplification and no detectable product is formed. The method works best when the mismatch is included in the 3'-most position of the oligonucleotide aligned with the polymorphism because this position is most destabilizing to elongation from the primer (see, e.g., WO 93/22456).

4. Direct-Sequencing

The direct analysis of the sequence of polymorphisms of the present invention can be accomplished using either the dideoxy chain termination method or the Maxam Gilbert method (see Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual* (2nd Ed., CSHP, New York 1989); Zyskind *et al.*, *Recombinant DNA Laboratory Manual*, (Acad. Press, 1988)).

5. Denaturing Gradient Gel Electrophoresis

Amplification products generated using the polymerase chain reaction can be analyzed by the use of denaturing gradient gel electrophoresis. Different alleles can be identified based on the different sequence-dependent melting properties and electrophoretic migration of DNA in solution. Erlich, ed., *PCR Technology, Principles and Applications for DNA Amplification*, (W.H. Freeman and Co, New York, 1992), Chapter 7.

6. Single-Strand Conformation Polymorphism Analysis

Alleles of target sequences can be differentiated using single-strand conformation polymorphism analysis, which identifies base differences by alteration in electrophoretic migration of single stranded PCR products, as described in Orita *et al.*, *Proc. Nat. Acad. Sci.* 86, 2766-2770 (1989). Amplified PCR products can be generated as described above, and heated or otherwise denatured, to form single stranded amplification products. Single-stranded nucleic acids may refold or form secondary structures which are partially dependent on the base sequence. The

different electrophoretic mobilities of single-stranded amplification products can be related to base-sequence differences between alleles of target sequences.

7. Single-Base Extension

An alternative method for identifying and analyzing polymorphisms is based on single-base extension (SBE) of a fluorescently-labeled primer coupled with fluorescence resonance energy transfer (FRET) between the label of the added base and the label of the primer. Typically, the method, such as that described by Chen *et al.*, (*PNAS 94*:10756-61 (1997), incorporated herein by reference) uses a locus-specific oligonucleotide primer labeled on the 5' terminus with 5-carboxyfluorescein (FAM). This labeled primer is designed so that the 3' end is immediately adjacent to the polymorphic site of interest. The labeled primer is hybridized to the locus, and single base extension of the labeled primer is performed with fluorescently labeled dideoxyribonucleotides (ddNTPs) in dye-terminator sequencing fashion, except that no deoxyribonucleotides are present. An increase in fluorescence of the added ddNTP in response to excitation at the wavelength of the labeled primer is used to infer the identity of the added nucleotide.

III. Methods of Use

After determining polymorphic form(s) present in an individual at one or more polymorphic sites, this information can be used in a number of methods.

A. Forensics

Determination of which polymorphic forms occupy a set of polymorphic sites in an individual identifies a set of polymorphic forms that distinguishes the individual. *See generally* National Research Council, *The Evaluation of Forensic DNA Evidence* (Eds. Pollard *et al.*, National Academy Press, DC, 1996). The more sites that are analyzed, the lower the probability that the set of polymorphic forms in one individual is the same as that in an unrelated individual. Preferably, if multiple sites are analyzed, the sites are unlinked. Thus, polymorphisms of the invention are often used in conjunction with polymorphisms in distal genes. Preferred polymorphisms for use in forensics are biallelic because the population frequencies

of two polymorphic forms can usually be determined with greater accuracy than those of multiple polymorphic forms at multi-allelic loci.

The capacity to identify a distinguishing or unique set of forensic markers in an individual is useful for forensic analysis. For example, one can determine whether a blood sample from a suspect matches a blood or other tissue sample from a crime scene by determining whether the set of polymorphic forms occupying selected polymorphic sites is the same in the suspect and the sample. If the set of polymorphic markers does not match between a suspect and a sample, it can be concluded (barring experimental error) that the suspect was not the source of the sample. If the set of markers does match, one can conclude that the DNA from the suspect is consistent with that found at the crime scene. If frequencies of the polymorphic forms at the loci tested have been determined (e.g., by analysis of a suitable population of individuals), one can perform a statistical analysis to determine the probability that a match of suspect and crime scene sample would occur by chance.

p(ID) is the probability that two random individuals have the same polymorphic or allelic form at a given polymorphic site. In biallelic loci, four genotypes are possible: AA, AB, BA, and BB. If alleles A and B occur in a haploid genome of the organism with frequencies x and y, the probability of each genotype in a diploid organism is (see WO 95/12607):

Homozygote: $p(AA)= x^2$

Homozygote: $p(BB)= y^2 = (1-x)^2$

Single Heterozygote: $p(AB)= p(BA)= xy = x(1-x)$

Both Heterozygotes: $p(AB+BA)= 2xy = 2x(1-x)$

The probability of identity at one locus (i.e, the probability that two individuals, picked at random from a population will have identical polymorphic forms at a given locus) is given by the equation:

$p(ID) = (x^2)^2 + (2xy)^2 + (y^2)^2$.

These calculations can be extended for any number of polymorphic forms at a given locus. For example, the probability of identity p(ID) for a 3-allele system

where the alleles have the frequencies in the population of x, y and z, respectively, is equal to the sum of the squares of the genotype frequencies:

$p(ID) = x^4 + (2xy)^2 + (2yz)^2 + (2xz)^2 + z^4 + y^4$

In a locus of n alleles, the appropriate binomial expansion is used to calculate p(ID) and p(exc).

The cumulative probability of identity (cum p(ID)) for each of multiple unlinked loci is determined by multiplying the probabilities provided by each locus.

cum p(ID) = p(ID1)p(ID2)p(ID3).... p(IDn)

The cumulative probability of non-identity for n loci (i.e. the probability that two random individuals will be different at 1 or more loci) is given by the equation:

cum p(nonID) = 1-cum p(ID).

If several polymorphic loci are tested, the cumulative probability of non-identity for random individuals becomes very high (e.g., one billion to one). Such probabilities can be taken into account together with other evidence in determining the guilt or innocence of the suspect.

B. Paternity Testing

The object of paternity testing is usually to determine whether a male is the father of a child. In most cases, the mother of the child is known and thus, the mother's contribution to the child's genotype can be traced. Paternity testing investigates whether the part of the child's genotype not attributable to the mother is consistent with that of the putative father. Paternity testing can be performed by analyzing sets of polymorphisms in the putative father and the child.

If the set of polymorphisms in the child attributable to the father does not match the set of polymorphisms of the putative father, it can be concluded, barring experimental error, that the putative father is not the real father. If the set of polymorphisms in the child attributable to the father does match the set of polymorphisms of the putative father, a statistical calculation can be performed to determine the probability of coincidental match.

The probability of parentage exclusion (representing the probability that a random male will have a polymorphic form at a given polymorphic site that makes him incompatible as the father) is given by the equation (see WO 95/12607):

p(exc) = xy(1-xy)

where x and y are the population frequencies of alleles A and B of a biallelic polymorphic site.

(At a triallelic site p(exc) = xy(1-xy) + yz(1- yz) + xz(1-xz)+ 3xyz(1-xyz))), where x, y and z and the respective population frequencies of alleles A, B and C).

The probability of non-exclusion is

p(non-exc) = 1-p(exc)

The cumulative probability of non-exclusion (representing the value obtained when n loci are used) is thus:

cum p(non-exc) = p(non-exc1)p(non-exc2)p(non-exc3).... p(non-excn)

The cumulative probability of exclusion for n loci (representing the probability that a random male will be excluded)

cum p(exc) = 1 - cum p(non-exc).

If several polymorphic loci are included in the analysis, the cumulative probability of exclusion of a random male is very high. This probability can be taken into account in assessing the liability of a putative father whose polymorphic marker set matches the child's polymorphic marker set attributable to his/her father.

C. Correlation of Polymorphisms with Phenotypic Traits

The polymorphisms of the invention may contribute to the phenotype of an organism in different ways. Some polymorphisms occur within a protein coding sequence and contribute to phenotype by affecting protein structure. The effect may be neutral, beneficial or detrimental, or both beneficial and detrimental, depending on the circumstances. For example, a heterozygous sickle cell mutation confers resistance to malaria, but a homozygous sickle cell mutation is usually lethal. Other polymorphisms occur in noncoding regions but may exert phenotypic effects indirectly via influence on replication, transcription, and translation. A single polymorphism may affect more than one phenotypic trait. Likewise, a single phenotypic trait may be affected by polymorphisms in different genes. Further, some polymorphisms predispose an individual to a distinct mutation that is causally related to a certain phenotype.

Phenotypic traits include diseases that have known but hitherto unmapped genetic components (e.g., agammaglobulimenia, diabetes insipidus, Lesch-Nyhan syndrome, muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial hypercholesterolemia, polycystic kidney disease, hereditary spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary hemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos syndrome, osteogenesis imperfecta, and acute intermittent porphyria). Phenotypic traits also include symptoms of, or susceptibility to, multifactorial diseases of which a component is or may be genetic, such as autoimmune diseases, inflammation, cancer, diseases of the nervous system, and infection by pathogenic microorganisms. Some examples of autoimmune diseases include rheumatoid arthritis, multiple sclerosis, diabetes (insulin-dependent and non-independent), systemic lupus erythematosus and Graves disease. Some examples of cancers include cancers of the bladder, brain, breast, colon, esophagus, kidney, leukemia, liver, lung, oral cavity, ovary, pancreas, prostate, skin, stomach and uterus. Phenotypic traits also include characteristics such as longevity, appearance (e.g., baldness, obesity), strength, speed, endurance, fertility, and susceptibility or receptivity to particular drugs or therapeutic treatments.

The correlation of one or more polymorphisms with phenotypic traits can be facilitated by knowledge of the gene product of the wild type (reference) gene. The genes in which cSNPs of the present invention have been identified are genes which have been previously sequenced and characterized in one of their allelic forms.

Correlation is performed for a population of individuals who have been tested for the presence or absence of a phenotypic trait of interest and for polymorphic markers sets. To perform such analysis, the presence or absence of a set of polymorphisms (i.e. a polymorphic set) is determined for a set of the individuals, some of whom exhibit a particular trait, and some of which exhibit lack of the trait. The alleles of each polymorphism of the set are then reviewed to determine whether the presence or absence of a particular allele is associated with the trait of interest. Correlation can be performed by standard statistical methods such as a κ-squared test and statistically significant correlations between polymorphic form(s) and phenotypic characteristics are noted. For example, it might be found that the presence of allele A1 at polymorphism A correlates with heart disease. As a further

example, it might be found that the combined presence of allele A1 at polymorphism A and allele B1 at polymorphism B correlates with increased milk production of a farm animal.

Such correlations can be exploited in several ways. In the case of a strong correlation between a set of one or more polymorphic forms and a disease for which treatment is available, detection of the polymorphic form set in a human or animal patient may justify immediate administration of treatment, or at least the institution of regular monitoring of the patient. Detection of a polymorphic form correlated with serious disease in a couple contemplating a family may also be valuable to the couple in their reproductive decisions. For example, the female partner might elect to undergo *in vitro* fertilization to avoid the possibility of transmitting such a polymorphism from her husband to her offspring. In the case of a weaker, but still statistically significant correlation between a polymorphic set and human disease, immediate therapeutic intervention or monitoring may not be justified. Nevertheless, the patient can be motivated to begin simple life-style changes (e.g., diet, exercise) that can be accomplished at little cost to the patient but confer potential benefits in reducing the risk of conditions to which the patient may have increased susceptibility by virtue of variant alleles. Identification of a polymorphic set in a patient correlated with enhanced receptiveness to one of several treatment regimes for a disease indicates that this treatment regime should be followed.

For animals and plants, correlations between characteristics and phenotype are useful for breeding for desired characteristics. For example, Beitz *et al.*, US 5,292,639 discuss use of bovine mitochondrial polymorphisms in a breeding program to improve milk production in cows. To evaluate the effect of mtDNA D-loop sequence polymorphism on milk production, each cow was assigned a value of 1 if variant or 0 if wildtype with respect to a prototypical mitochondrial DNA sequence at each of 17 locations considered. Each production trait was analyzed individually with the following animal model:

$$Y_{ijkpn} = \mu + YS_i + P_j + X_k + \beta_1 + \ldots \beta_{17} + PE_n + a_n + e_p$$

where Y_{ijknp} is the milk, fat, fat percentage, SNF, SNF percentage, energy concentration, or lactation energy record; μ is an overall mean; YS_i is the effect common to all cows calving in year-season; X_k is the effect common to cows in

either the high or average selection line; β_1 to β_{17} are the binomial regressions of production record on mtDNA D-loop sequence polymorphisms; PE_n is permanent environmental effect common to all records of cow n; a_n is effect of animal n and is composed of the additive genetic contribution of sire and dam breeding values and a Mendelian sampling effect; and e_p is a random residual. It was found that eleven of seventeen polymorphisms tested influenced at least one production trait. Bovines having the best polymorphic forms for milk production at these eleven loci are used as parents for breeding the next generation of the herd.

D. Genetic Mapping of Phenotypic Traits

The previous section concerns identifying correlations between phenotypic traits and polymorphisms that directly or indirectly contribute to those traits. The present section describes identification of a physical linkage between a genetic locus associated with a trait of interest and polymorphic markers that are not associated with the trait, but are in physical proximity with the genetic locus responsible for the trait and co-segregate with it. Such analysis is useful for mapping a genetic locus associated with a phenotypic trait to a chromosomal position, and thereby cloning gene(s) responsible for the trait. *See* Lander *et al.*, *Proc. Natl. Acad. Sci. (USA)* 83, 7353-7357 (1986); Lander *et al.*, *Proc. Natl. Acad. Sci. (USA)* 84, 2363-2367 (1987); Donis-Keller *et al.*, *Cell* 51, 319-337 (1987); Lander *et al.*, *Genetics* 121, 185-199 (1989)). Genes localized by linkage can be cloned by a process known as directional cloning. *See* Wainwright, *Med. J. Australia* 159, 170-174 (1993); Collins, *Nature Genetics* 1, 3-6 (1992).

Linkage studies are typically performed on members of a family. Available members of the family are characterized for the presence or absence of a phenotypic trait and for a set of polymorphic markers. The distribution of polymorphic markers in an informative meiosis is then analyzed to determine which polymorphic markers co-segregate with a phenotypic trait. *See, e.g.*, Kerem *et al.*, *Science* 245, 1073-1080 (1989); Monaco *et al.*, *Nature* 316, 842 (1985); Yamoka *et al.*, *Neurology* 40, 222-226 (1990); Rossiter *et al.*, *FASEB Journal* 5, 21-27 (1991).

Linkage is analyzed by calculation of LOD (log of the odds) values. A lod value is the relative likelihood of obtaining observed segregation data for a marker

and a genetic locus when the two are located at a recombination fraction θ, versus the situation in which the two are not linked, and thus segregating independently (Thompson & Thompson, *Genetics in Medicine* (5th ed, W.B. Saunders Company, Philadelphia, 1991); Strachan, "Mapping the human genome" in *The Human Genome* (BIOS Scientific Publishers Ltd, Oxford), Chapter 4). A series of likelihood ratios are calculated at various recombination fractions (θ), ranging from θ = 0.0 (coincident loci) to θ = 0.50 (unlinked). Thus, the likelihood at a given value of θ is: probability of data if loci linked at θ to probability of data if loci unlinked. The computed likelihoods are usually expressed as the $\log_{10}$ of this ratio (i.e., a lod score). For example, a lod score of 3 indicates 1000:1 odds against an apparent observed linkage being a coincidence. The use of logarithms allows data collected from different families to be combined by simple addition. Computer programs are available for the calculation of lod scores for differing values of θ (e.g., LIPED, MLINK (Lathrop, *Proc. Nat. Acad. Sci. (USA)* 81, 3443-3446 (1984)). For any particular lod score, a recombination fraction may be determined from mathematical tables. *See* Smith *et al.*, *Mathematical tables for research workers in human genetics* (Churchill, London, 1961); Smith, *Ann. Hum. Genet.* 32, 127-150 (1968). The value of θ at which the lod score is the highest is considered to be the best estimate of the recombination fraction.

Positive lod score values suggest that the two loci are linked, whereas negative values suggest that linkage is less likely (at that value of θ) than the possibility that the two loci are unlinked. By convention, a combined lod score of +3 or greater (equivalent to greater than 1000:1 odds in favor of linkage) is considered definitive evidence that two loci are linked. Similarly, by convention, a negative lod score of -2 or less is taken as definitive evidence against linkage of the two loci being compared. Negative linkage data are useful in excluding a chromosome or a segment thereof from consideration. The search focuses on the remaining non-excluded chromosomal locations.

IV. Modified Polypeptides and Gene Sequences

The invention further provides variant forms of nucleic acids and corresponding proteins. The nucleic acids comprise one of the sequences described

in the Table, column 5, in which the polymorphic position is occupied by one of the alternative bases for that position. Some nucleic acids encode full-length variant forms of proteins. Similarly, variant proteins have the prototypical amino acid sequences encoded by nucleic acid sequences shown in the Table, column 5, (read so as to be in-frame with the full-length coding sequence of which it is a component) except at an amino acid encoded by a codon including one of the polymorphic positions shown in the Table. That position is occupied by the amino acid coded by the corresponding codon in any of the alternative forms shown in the Table.

Variant genes can be expressed in an expression vector in which a variant gene is operably linked to a native or other promoter. Usually, the promoter is a eukaryotic promoter for expression in a mammalian cell. The transcription regulation sequences typically include a heterologous promoter and optionally an enhancer which is recognized by the host. The selection of an appropriate promoter, for example trp, lac, phage promoters, glycolytic enzyme promoters and tRNA promoters, depends on the host selected. Commercially available expression vectors can be used. Vectors can include host-recognized replication systems, amplifiable genes, selectable markers, host sequences useful for insertion into the host genome, and the like.

The means of introducing the expression construct into a host cell varies depending upon the particular construction and the target host. Suitable means include fusion, conjugation, transfection, transduction, electroporation or injection, as described in Sambrook, *supra*. A wide variety of host cells can be employed for expression of the variant gene, both prokaryotic and eukaryotic. Suitable host cells include bacteria such as *E. coli*, yeast, filamentous fungi, insect cells, mammalian cells, typically immortalized, *e.g.*, mouse, CHO, human and monkey cell lines and derivatives thereof. Preferred host cells are able to process the variant gene product to produce an appropriate mature polypeptide. Processing includes glycosylation, ubiquitination, disulfide bond formation, general post-translational modification, and the like. As used herein, "gene product" includes mRNA, peptide and protein products.

The protein may be isolated by conventional means of protein biochemistry and purification to obtain a substantially pure product, *i.e.*, 80, 95 or 99% free of cell

component contaminants, as described in Jacoby, *Methods in Enzymology* Volume 104, Academic Press, New York (1984); Scopes, *Protein Purification, Principles and Practice*, 2nd Edition, Springer-Verlag, New York (1987); and Deutscher (ed), *Guide to Protein Purification, Methods in Enzymology*, Vol. 182 (1990). If the protein is secreted, it can be isolated from the supernatant in which the host cell is grown. If not secreted, the protein can be isolated from a lysate of the host cells.

The invention further provides transgenic nonhuman animals capable of expressing an exogenous variant gene and/or having one or both alleles of an endogenous variant gene inactivated. Expression of an exogenous variant gene is usually achieved by operably linking the gene to a promoter and optionally an enhancer, and microinjecting the construct into a zygote. *See* Hogan *et al.*, "Manipulating the Mouse Embryo, A Laboratory Manual," Cold Spring Harbor Laboratory. Inactivation of endogenous variant genes can be achieved by forming a transgene in which a cloned variant gene is inactivated by insertion of a positive selection marker. *See* Capecchi, *Science* 244, 1288-1292 (1989). The transgene is then introduced into an embryonic stem cell, where it undergoes homologous recombination with an endogenous variant gene. Mice and other rodents are preferred animals. Such animals provide useful drug screening systems.

In addition to substantially full-length polypeptides expressed by variant genes, the present invention includes biologically active fragments of the polypeptides, or analogs thereof, including organic molecules which simulate the interactions of the peptides. Biologically active fragments include any portion of the full-length polypeptide which confers a biological function on the variant gene product, including ligand binding, and antibody binding. Ligand binding includes binding by nucleic acids, proteins or polypeptides, small biologically active molecules, or large cellular structures.

Polyclonal and/or monoclonal antibodies that specifically bind to variant gene products but not to corresponding prototypical gene products are also provided. Antibodies can be made by injecting mice or other animals with the variant gene product or synthetic peptide fragments thereof. Monoclonal antibodies are screened as are described, for example, in Harlow & Lane, *Antibodies, A Laboratory Manual*, Cold Spring Harbor Press, New York (1988); Goding, *Monoclonal antibodies,*

Principles and Practice (2d ed.) Academic Press, New York (1986). Monoclonal antibodies are tested for specific immunoreactivity with a variant gene product and lack of immunoreactivity to the corresponding prototypical gene product. These antibodies are useful in diagnostic assays for detection of the variant form, or as an active ingredient in a pharmaceutical composition.

V. Kits

The invention further provides kits comprising at least one allele-specific oligonucleotide as described herein. Often, the kits contain one or more pairs of allele-specific oligonucleotides hybridizing to different forms of a polymorphism. In some kits, the allele-specific oligonucleotides are provided immobilized to a substrate. For example, the same substrate can comprise allele-specific oligonucleotide probes for detecting at least 10, 100 or all of the polymorphisms shown in the Table. Optional additional components of the kit include, for example, restriction enzymes, reverse-transcriptase or polymerase, the substrate nucleoside triphosphates, means used to label (for example, an avidin-enzyme conjugate and enzyme substrate and chromogen if the label is biotin), and the appropriate buffers for reverse transcription, PCR, or hybridization reactions. Usually, the kit also contains instructions for carrying out the methods.

The thrombospondins are a family of extracellular matrix (ECM) glycoproteins that modulate many cell behaviors including adhesion, migration, and proliferation. Thrombospondins (also known as thrombin sensitive proteins or TSPs) are large molecular weight glycoproteins composed of three identical disulfide-linked polypeptide chains. TSPs are stored in the alpha-granules of platelets and secreted by a variety of mesenchymal and epithelial cells (Majack *et al.*, *Cell Membrane 3*:57-77 (1987)). Platelets secrete TSPs when activated in the blood by such physiological agonists such as thrombin. TSPs have lectin properties and a broad function in the regulation of fibrinolysis and as a component of the ECM, and are one of a group of ECM proteins which have adhesive properties. TSPs bind to fibronectin and fibrinogen (Lahav *et al.*, *Eur J Biochem 145*:151-6 (1984)), and these proteins are known to be involved in platelet adhesion to substratum and platelet aggregation (Leung, *J Clin Invest 74*:1764-1772 (1986)).

Recent work has implicated TSPs in response of cells to growth factors. Submitogenic doses of PDGF induce a rapid but transitory, increase in TSP synthesis and secretion by rat aortic smooth muscle cells (Majack *et al.*, *J Biol Chem 101*:1059-70 (1985)). PDGF responsiveness to TSP synthesis in glial cells has also been shown (Asch *et al.*, *Proc Natl Acad Sci 83*:2904-8 (1986)). TSP mRNA levels rise rapidly in response to PDGF (Majack *et al.*, *J Biol Chem 262*:8821-5 (1987)). TSPs act synergistically with epidermal growth factor to increase DNA synthesis in smooth muscle cells (Majack *et al.*, *Proc Natl Acad Sci 83*:9050-4 (1986)), and monoclonal antibodies to TSPs inhibit smooth muscle cell proliferation (Majack *et al.*, *J Biol Chem 106*:415-22 (1988)). TSPs modulate local adhesions in endothelial cells, and TSPs, particularly TSP-1 primarily derived from platelet granules, are known to be an important activator of transforming growth factor beta-1 (TGFB-1) (Crawford *et al.*, *Cell 93*:1159 (1998)) and appear to be a potential link between platelet-thrombosis and development of atherosclerosis.

To determine pivotal genes associated with premature coronary artery disease, we analyzed DNA from 347 patients with MI or coronary revascularization before age 40 (men) or 45 (women) and 422 general population controls. Cases were drawn (one per family) from a retrospective collection of sibling pairs with premature CAD. Controls were ascertained through random-digit dialing. Both cases and controls were Caucasian. A complete database of phenotypic and laboratory variables for the affected patients afforded logistic regression to control for age, diabetes, body mass index, gender.

Thrombospondin (TSP) 4 and 1 emerged as important SNPs associated with premature CAD and MI. For CAD, 148 of 347 patients carried at least one copy of the TSP-4 variant compared with 142 of 422 control subjects; adjusted odds ratio 1.47, p=0.01. For premature MI, the association was even stronger: 91 of 187 cases vs. 142 of 422 controls had the variant; adjusted odds ratio 2.08, p=0.0003. The TSP-1 SNP was rare. Nonetheless, homozygosity for the variant allele gave an adjusted odds ratio of 9.5, p=.04.

Specific reference nucleotide (SEQ ID NO: 1) and amino acid (SEQ ID NO: 2) sequences for TSP-1 are shown in Figs. 1A-1D. Specific reference nucleotide (SEQ ID NO: 3) and amino acid (SEQ ID NO: 4) sequences for TSP-4 are shown in

Figs. 2A-2C. It is understood that the invention is not limited by these exemplified reference sequences, as variants of these sequences which differ at locations other than the SNP sites identified herein can also be utilized. The skilled artisan can readily determine the SNP sites in these other reference sequences which correspond to the SNP sites identified herein by aligning the sequence of interest with the reference sequences specifically disclosed herein, and programs for performing such alignments are commercially available. For example, the ALIGN program in the GCG software package can be used, utilizing a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4, for example.

Two SNPs have been specifically studied as described herein. The first (G334u4) is a change from A (reference nucleotide) to G (alternate or variant nucleotide) at nucleotide position 2210 of the nucleic acid sequence of TSP-1 (Figs. 1A-1D), resulting in a missense amino acid mutation from asparagine (reference) to serine (alternate) at amino acid 700. The second SNP (G355u2) is a change from G (reference) to C (alternate) at nucleotide position 1186 of the nucleic acid sequence of TSP-4 (Figs. 2A-2C), resulting in a missense amino acid alteration from alanine (reference) to proline (alternate) at amino acid 387. With respect to the G355u2 SNP, individuals with CAD carried at least one copy of the variant "C" allele more frequently than control individuals (43% as compared with 34%). With respect to the G355u2 SNP, individuals with MI carried at least one copy of the variant "C" allele more frequently than control individuals (49% as compared with 34%). With respect to the G334u4 SNP, individuals with CAD carried two copies of the variant "G" allele more frequently than control individuals (1.7% as compared with 0.2%). With respect to the G334u4 SNP, individuals with MI carried two copies of the variant "G" allele more frequently than control individuals (2% as compared with 0.2%).

As used herein, the term "polymorphism" refers to the occurrence of two or more genetically determined alternative sequences or alleles in a population. A polymorphic marker or site is the locus at which divergence occurs. Preferred markers have at least two alleles, each occurring at frequency of greater than 1%, and more preferably greater than 10% or 20% of a selected population. A

polymorphic locus may be as small as one base pair, in which case it is referred to as a single nucleotide polymorphism (SNP).

Thus, the invention relates to a method for predicting the likelihood that an individual will have a vascular disease, or for aiding in the diagnosis of a vascular disease, or predicting the likelihood of having altered symptomology associated with a vascular disease, comprising the steps of obtaining a DNA sample from an individual to be assessed and determining the nucleotide present at one or more of nucleotide positions 2210 of the TSP-1 gene or 1186 of the TSP-4 gene. In a preferred embodiment, the nucleotides present at both of these nucleotide positions are determined. In one embodiment the TSP-1 gene has the nucleotide sequence of SEQ ID NO: 1 and the TSP-4 gene has the nucleotide sequence of SEQ ID NO: 3. The presence of one or more of a G (the variant nucleotide) at position 2210 of SEQ ID NO: 1 or a C (the variant nucleotide) at position 1186 of SEQ ID NO: 1186 indicates that the individual has a greater likelihood of having a vascular disease, or a greater likelihood of having severe symptomology associated with a vascular disease, than if that individual had the reference nucleotide at one or more of these positions. Conversely, the presence of one or more of an A (the reference nucleotide) at position 2210 of SEQ ID NO: 1 or a G (the reference nucleotide) at position 1186 of SEQ ID NO: 3 indicates that the individual has a reduced likelihood of having a vascular disease or a likelihood of having reduced symptomology associated with a vascular disease than if that individual had the variant nucleotide at one or more of these positions.

In a particular embodiment, the individual is an individual at risk for development of a vascular disease. In another embodiment the individual exhibits clinical symptomology associated with a vascular disease. In one embodiment, the individual has been clinically diagnosed as having a vascular disease. Vascular diseases include, but are not limited to, atherosclerosis, coronary heart disease, myocardial infarction (MI), stroke, peripheral vascular diseases, venous thromboembolism and pulmonary embolism. In preferred embodiments, the vascular disease is CAD or MI.

The genetic material to be assessed can be obtained from any nucleated cell from the individual. For assay of genomic DNA, virtually any biological sample

(other than pure red blood cells) is suitable. For example, convenient tissue samples include whole blood, semen, saliva, tears, urine, fecal material, sweat, skin and hair. For assay of cDNA or mRNA, the tissue sample must be obtained from a tissue or organ in which the target nucleic acid is expressed.

Many of the methods described herein require amplification of DNA from target samples. This can be accomplished by e.g., PCR. *See generally PCR Technology: Principles and Applications for DNA Amplification* (ed. H.A. Erlich, Freeman Press, NY, NY, 1992); *PCR Protocols: A Guide to Methods and Applications* (eds. Innis, *et al.*, Academic Press, San Diego, CA, 1990); Mattila *et al.*, *Nucleic Acids Res.* 19, 4967 (1991); Eckert *et al.*, *PCR Methods and Applications* 1, 17 (1991); *PCR* (eds. McPherson *et al.*, IRL Press, Oxford); and U.S. Patent 4,683,202.

Other suitable amplification methods include the ligase chain reaction (LCR) (see Wu and Wallace, *Genomics* 4, 560 (1989), Landegren *et al.*, *Science* 241, 1077 (1988), transcription amplification (Kwoh *et al.*, *Proc. Natl. Acad. Sci. USA* 86, 1173 (1989)), and self-sustained sequence replication (Guatelli *et al.*, *Proc. Nat. Acad. Sci. USA,* 87, 1874 (1990)) and nucleic acid based sequence amplification (NASBA). The latter two amplification methods involve isothermal reactions based on isothermal transcription, which produce both single stranded RNA (ssRNA) and double stranded DNA (dsDNA) as the amplification products in a ratio of about 30 or 100 to 1, respectively.

The nucleotide which occupies the polymorphic site of interest (e.g., nucleotide position 2210 in TSP-1 and/or nucleotide position 1186 in TSP-4) can be identified by a variety of methods, such as Southern analysis of genomic DNA; direct mutation analysis by restriction enzyme digestion; Northern analysis of RNA; denaturing high pressure liquid chromatography (DHPLC); gene isolation and sequencing; hybridization of an allele-specific oligonucleotide with amplified gene products; single base extension (SBE). In a preferred embodiment, determination of the allelic form of TSP is carried out using SBE-FRET methods as described herein, or using chip-based oligonucleotide arrays as described herein.

The invention also relates to a method for predicting the likelihood that an individual will have a vascular disease, or for aiding in the diagnosis of a vascular

disease, or predicting the likelihood of having altered symptomology associated with a vascular disease, comprising the steps of obtaining a biological sample comprising TSP-1 and/or TSP-4 protein or relevant portion thereof from an individual to be assessed and determining the amino acid present at one or more of amino acid positions 700 of the TSP-1 gene product (e.g., as exemplified by SEQ ID NO: 2) or 387 of the TSP-4 gene product (e.g., as exemplified by SEQ ID NO: 4). In a preferred embodiment, the amino acids present at both of these amino acid positions are determined. As used herein, the term "relevant portion" of the TSP-1 and TSP-4 proteins is intended to encompass any portion of the protein which comprises the polymorphic amino acid positions. The presence of one or more of a serine (the variant amino acid) at position 700 of SEQ ID NO: 2, or a proline (the variant amino acid) at position 387 of SEQ ID NO: 4 indicates that the individual has a greater likelihood of having a vascular disease, or a greater likelihood of having severe symptomology associated with a vascular disease, than if that individual had the reference amino acid at one or more of these positions. Conversely, the presence of one or more of an asparagine (the reference amino acid) at position 700 of SEQ ID NO: 2, or an alanine (the reference amino acid) at position 387 of SEQ I D NO: 4 indicates that the individual has a reduced likelihood of having a vascular disease or a likelihood of having reduced symptomology associated with a vascular disease, than if that individual had the varaint amino acid at one or more of these positions.

In a particular embodiment, the individual is an individual at risk for development of a vascular disease. In another embodiment the individual exhibits clinical symptomology associated with a vascular disease. In one embodiment, the individual has been clinically diagnosed as having a vascular disease.

In this embodiment of the invention, the biological sample contains protein molecules from the test subject. *In vitro* techniques for detection of protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. Furthermore, *in vivo* techniques for detection of protein include introducing into a subject a labeled anti-protein antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques. Polyclonal and/or monoclonal antibodies that specifically bind to variant gene

products but not to corresponding reference gene products, and vice versa, are also provided. Antibodies can be made by injecting mice or other animals with the variant gene product or synthetic peptide fragments thereof comprising the variant portion. Monoclonal antibodies are screened as are described, for example, in Harlow & Lane, *Antibodies, A Laboratory Manual*, Cold Spring Harbor Press, New York (1988); Goding, *Monoclonal antibodies, Principles and Practice* (2d ed.) Academic Press, New York (1986). Monoclonal antibodies are tested for specific immunoreactivity with a variant gene product and lack of immunoreactivity to the corresponding prototypical gene product. These antibodies are useful in diagnostic assays for detection of the variant form, or as an active ingredient in a pharmaceutical composition.

The polymorphisms of the invention may be associated with vascular disease in different ways. The polymorphisms may exert phenotypic effects indirectly via influence on replication, transcription, and translation. Additionally, the described polymorphisms may predispose an individual to a distinct mutation that is causally related to a certain phenotype, such as susceptibility or resistance to vascular disease and related disorders. The discovery of the polymorphisms and their correlation with CAD and MI facilitates biochemical analysis of the variant and reference forms and the development of assays to characterize the variant and reference forms and to screen for pharmaceutical agents that interact directly with one or another form of the protein.

Alternatively, these particular polymorphisms may belong to a group of two or more polymorphisms in the TSP gene(s) which contributes to the presence, absence or severity of vascular disease. An assessment of other polymorphisms within the TSP gene(s) can be undertaken, and the separate and combined effects of these polymorphisms, as well as alternations in other, distinct genes, on the vascular disease phenotype can be assessed.

Correlation between a particular phenotype, e.g., the CAD or MI phenotype, and the presence or absence of a particular allele is performed for a population of individuals who have been tested for the presence or absence of the phenotype. Correlation can be performed by standard statistical methods such as a Chi-squared test and statistically significant correlations between polymorphic form(s) and

phenotypic characteristics are noted. This correlation can be exploited in several ways. In the case of a strong correlation between a particular polymorphic form, e.g., the variant allele for TSP-1 and/or TSP-4, and a disease for which treatment is available, detection of the polymorphic form in an individual may justify immediate administration of treatment, or at least the institution of regular monitoring of the individual. Detection of a polymorphic form correlated with a disorder in a couple contemplating a family may also be valuable to the couple in their reproductive decisions. For example, the female partner might elect to undergo *in vitro* fertilization to avoid the possibility of transmitting such a polymorphism from her husband to her offspring. In the case of a weaker, but still statistically significant correlation between a polymorphic form and a particular disorder, immediate therapeutic intervention or monitoring may not be justified. Nevertheless, the individual can be motivated to begin simple life-style changes (e.g., diet modification, therapy or counseling) that can be accomplished at little cost to the individual but confer potential benefits in reducing the risk of conditions to which the individual may have increased susceptibility by virtue of the particular allele. Furthermore, identification of a polymorphic form correlated with enhanced receptiveness to one of several treatment regimes for a disorder indicates that this treatment regimen should be followed for the individual in question.

Furthermore, it may be possible to identify a physical linkage between a genetic locus associated with a trait of interest (e.g., CAD or MI) and polymorphic markers that are or are not associated with the trait, but are in physical proximity with the genetic locus responsible for the trait and co-segregate with it. Such analysis is useful for mapping a genetic locus associated with a phenotypic trait to a chromosomal position, and thereby cloning gene(s) responsible for the trait. *See* Lander *et al.*, *Proc. Natl. Acad. Sci. (USA)* 83, 7353-7357 (1986); Lander *et al.*, *Proc. Natl. Acad. Sci. (USA)* 84, 2363-2367 (1987); Donis-Keller *et al.*, *Cell* 51, 319-337 (1987); Lander *et al.*, *Genetics* 121, 185-199 (1989)). Genes localized by linkage can be cloned by a process known as directional cloning. *See* Wainwright, *Med. J. Australia* 159, 170-174 (1993); Collins, *Nature Genetics* 1, 3-6 (1992). Linkage studies are discussed in more detail above.

In another embodiment, the invention relates to pharmaceutical compositions comprising a reference TSP-1 and/or TSP-4 gene or gene product for use in the treatment of vascular disease, e.g., CAD and MI. As used herein, a reference TSP gene product is intended to mean gene products which are encoded by the reference allele of the TSP gene. In addition to substantially full-length polypeptides expressed by the genes, the present invention includes biologically active fragments of the polypeptides, or analogs thereof, including organic molecules which simulate the interactions of the peptides. Biologically active fragments include any portion of the full-length polypeptide which confers a biological function on the variant gene product, including ligand binding, and antibody binding. Ligand binding includes binding by nucleic acids, proteins or polypeptides, small biologically active molecules, or large cellular structures.

For instance, the polypeptide or protein, or fragment thereof, of the present invention can be formulated with a physiologically acceptable medium to prepare a pharmaceutical composition. The particular physiological medium may include, but is not limited to, water, buffered saline, polyols (e.g., glycerol, propylene glycol, liquid polyethylene glycol) and dextrose solutions. The optimum concentration of the active ingredient(s) in the chosen medium can be determined empirically, according to procedures well known to medicinal chemists, and will depend on the ultimate pharmaceutical formulation desired. Methods of introduction of exogenous peptides at the site of treatment include, but are not limited to, intradermal, intramuscular, intraperitoneal, intravenous, subcutaneous, oral and intranasal. Other suitable methods of introduction can also include rechargeable or biodegradable devices and slow release polymeric devices. The pharmaceutical compositions of this invention can also be administered as part of a combinatorial therapy with other agents and treatment regimens.

The invention further pertains to compositions, e.g., vectors, comprising a nucleotide sequence encoding reference or variant TSP-1 and/or TSP-4 gene products. For example, reference genes can be expressed in an expression vector in which a reference gene is operably linked to a native or other promoter. Usually, the promoter is a eukaryotic promoter for expression in a mammalian cell. The transcription regulation sequences typically include a heterologous promoter and

optionally an enhancer which is recognized by the host. The selection of an appropriate promoter, for example trp, lac, phage promoters, glycolytic enzyme promoters and tRNA promoters, depends on the host selected. Commercially available expression vectors can be used. Vectors can include host-recognized replication systems, amplifiable genes, selectable markers, host sequences useful for insertion into the host genome, and the like.

The means of introducing the expression construct into a host cell varies depending upon the particular construction and the target host. Suitable means include fusion, conjugation, transfection, transduction, electroporation or injection, as described in Sambrook, *supra*. A wide variety of host cells can be employed for expression of the variant gene, both prokaryotic and eukaryotic. Suitable host cells include bacteria such as *E. coli*, yeast, filamentous fungi, insect cells, mammalian cells, typically immortalized, *e.g.*, mouse, CHO, human and monkey cell lines and derivatives thereof. Preferred host cells are able to process the variant gene product to produce an appropriate mature polypeptide. Processing includes glycosylation, ubiquitination, disulfide bond formation, general post-translational modification, and the like.

It is also contemplated that cells can be engineered to express the reference allele of the invention by gene therapy methods. For example, DNA encoding the reference TSP gene product, or an active fragment or derivative thereof, can be introduced into an expression vector, such as a viral vector, and the vector can be introduced into appropriate cells in an animal. In such a method, the cell population can be engineered to inducibly or constitutively express active reference TSP gene product. In a preferred embodiment, the vector is delivered to the bone marrow, for example as described in Corey *et al.* (*Science 244*:1275-1281 (1989)).

The invention further relates to the use of compositions (i.e., agonists) which enhance or increase the activity of the reference (or variant) TSP (e.g., TSP-1 or TSP-4) gene product, or a functional portion thereof, for use in the treatment of vascular disease. The invention also relates to the use of compositions (i.e., antagonists) which reduce or decrease the activity of the variant (or reference) TSP (e.g., TSP-1 or TSP-4) gene product, or a functional portion thereof, for use in the treatment of vascular disease.

The invention also relates to constructs which comprise a vector into which a sequence of the invention has been inserted in a sense or antisense orientation. For example, a vector comprising a nucleotide sequence which is antisense to the variant TSP-1 or TSP-4 allele may be used as an antagonist of the activity of the TSP-1 or TSP-4 variant allele. Alternatively, a vector comprising a nucleotide sequence of the TSP-1 or TSP-4 reference allele may be used therapeutically to treat vascular diseases. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors, expression vectors, are capable of directing the expression of genes to which they are operably linked. In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids (vectors). However, the invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses) that serve equivalent functions.

Preferred recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell. This means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operably linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to include promoters,

enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel, *Gene Expression Technology: Methods in Enzymology 185*, Academic Press, San Diego, CA (1990). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cell and those which direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc.

The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein . The recombinant expression vectors of the invention can be designed for expression of a polypeptide of the invention in prokaryotic or eukaryotic cells, *e.g.*, bacterial cells such as *E. coli*, insect cells (using baculovirus expression vectors), yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, *supra*. Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but also to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein. A host cell can be any prokaryotic or eukaryotic cell. For example, a nucleic acid of the invention can be expressed in bacterial cells (*e.g.*, *E. coli*), insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of

art-recognized techniques for introducing foreign nucleic acid (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.* (*supra*), and other laboratory manuals.

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) a polypeptide of the invention. Accordingly, the invention further provides methods for producing a polypeptide using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of the invention (into which a recombinant expression vector encoding a polypeptide of the invention has been introduced) in a suitable medium such that the polypeptide is produced. In another embodiment, the method further comprises isolating the polypeptide from the medium or the host cell.

The host cells of the invention can also be used to produce nonhuman transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which a nucleic acid of the invention has been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous nucleotide sequences have been introduced into their genome or homologous recombinant animals in which endogenous nucleotide sequences have been altered. Such animals are useful for studying the function and/or activity of the nucleotide sequence and polypeptide encoded by the sequence and for identifying and/or evaluating modulators of their activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, etc. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, an "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous gene has been altered by homologous

recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, *e.g.*, an embryonic cell of the animal, prior to development of the animal.

A transgenic animal of the invention can be created by introducing a nucleic acid of the invention into the male pronuclei of a fertilized oocyte, *e.g.*, by microinjection, retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. The sequence can be introduced as a transgene into the genome of a non-human animal. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably linked to the transgene to direct expression of a polypeptide in particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, U.S. Patent No. 4,873,191 and in Hogan, *Manipulating the Mouse Embryo* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the transgene in its genome and/or expression of mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene encoding the transgene can further be bred to other transgenic animals carrying other transgenes.

The invention also relates to the use of the variant and reference gene products to guide efforts to identify the causative mutation for vascular diseases or to identify or synthesize agents useful in the treatment of vascular diseases, e.g., CAD and MI. Amino acids that are essential for function can be identified by methods known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham *et al.*, *Science, 244*:1081-1085 (1989)). The latter procedure introduces single alanine mutations at every residue in the molecule. The resulting mutant molecules are then tested for biological activity *in vitro*, or *in vitro* activity. Sites that are critical for polypeptide activity can also be determined by structural analysis such as crystallization, nuclear magnetic resonance or photoaffinity labeling

(Smith *et al.*, *J. Mol. Biol.*, *224*:899-904 (1992); de Vos *et al. Science*, *255*:306-312 (1992)).

Another aspect of the invention pertains to monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of proteins of the invention in clinical trials. An exemplary method for detecting the presence or absence of proteins or nucleic acids of the invention in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting the protein, or nucleic acid (*e.g.*, mRNA, genomic DNA) that encodes the protein, such that the presence of the protein or nucleic acid is detected in the biological sample. A preferred agent for detecting mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to mRNA or genomic DNA sequences described herein, preferably in an allele-specific manner. The nucleic acid probe can be, for example, a full-length nucleic acid, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to appropriate mRNA or genomic DNA. Other suitable probes for use in the diagnostic assays of the invention are described herein.

The invention also encompasses kits for detecting the presence of proteins or nucleic acid molecules of the invention in a biological sample. For example, the kit can comprise a labeled compound or agent (e.g., nucleic acid probe) capable of detecting protein or mRNA in a biological sample; means for determining the amount of protein or mRNA in the sample; and means for comparing the amount of protein or mRNA in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect protein or nucleic acid.

The following Examples are offered for the purpose of illustrating the present invention and are not to be construed to limit the scope of this invention. The teachings of all references cited herein are hereby incorporated herein by reference.

EXAMPLES

Identification of Single Nucleotide Polymorphisms

The polymorphisms shown in the Table were identified by resequencing of target sequences from individuals of diverse ethnic and geographic backgrounds by hybridization to probes immobilized to microfabricated arrays. The strategy and principles for design and use of such arrays are generally described in WO 95/11995.

A typical probe array used in this analysis has two groups of four sets of probes that respectively tile both strands of a reference sequence. A first probe set comprises a plurality of probes exhibiting perfect complementarily with one of the reference sequences. Each probe in the first probe set has an interrogation position that corresponds to a nucleotide in the reference sequence. That is, the interrogation position is aligned with the corresponding nucleotide in the reference sequence, when the probe and reference sequence are aligned to maximize complementarily between the two. For each probe in the first set, there are three corresponding probes from three additional probe sets. Thus, there are four probes corresponding to each nucleotide in the reference sequence. The probes from the three additional probe sets are identical to the corresponding probe from the first probe set except at the interrogation position, which occurs in the same position in each of the four corresponding probes from the four probe sets, and is occupied by a different nucleotide in the four probe sets. In the present analysis, probes were 25 nucleotides long. Arrays tiled for multiple different references sequences were included on the same substrate.

Publicly available sequences for a given gene were assembled into Gap4 (http://www.biozentrum.unibas.ch/~biocomp/staden/Overview.html). PCR primers covering each exon were designed using Primer 3 (http://www-genome.wi.mit.edu/cgi-bin/primer/primer3.cgi). Primers were not designed in regions where there were sequence discrepancies between reads. Genomic DNA was amplified in at least 50 individuals using 2.5 pmol each primer, 1.5 mM $MgCl_2$, 100 μM dNTPs, 0.75 μM AmpliTaq GOLD polymerase, and 19 ng DNA in a 15 μl reaction. Reactions were assembled using a PACKARD MultiPROBE robotic pipetting station and then put in MJ 96-well tetrad thermocyclers (96°C for 10

minutes, followed by 35 cycles of 96°C for 30 seconds, 59°C for 2 minutes, and 72°C for 2 minutes). A subset of the PCR assays for each individual were run on 3% NuSieve gels in 0.5X TBE to confirm that the reaction worked.

For a given DNA, 5 µl (about 50 ng) of each PCR or RT-PCR product were pooled (Final volume = 150-200 µl). The products were purified using QiaQuick PCR purification from Qiagen. The samples were eluted once in 35 µl sterile water and 4 µl 10X One-Phor-All buffer (Pharmacia). The pooled samples were digested with 0.2 µ DNaseI (Promega)for 10 minutes at 37°C and then labeled with 0.5 nmols biotin-N6-ddATP and 15 µ Terminal Transferase (GibcoBRL Life Technology) for 60 minutes at 37°C. Both fragmentation and labeling reactions were terminated by incubating the pooled sample for 15 minutes at 100°C.

Low-density DNA chips (Affymetrix,CA) were hybridized following the manufacturer's instructions. Briefly, the hybridization cocktail consisted of 3M TMACl, 10 mM Tris pH 7.8, 0.01% Triton X-100, 100 mg/ml herring sperm DNA (Gibco BRL), 200 pM control biotin-labeled oligo. The processed PCR products were denatured for 7 minutes at 100°C and then added to prewarmed (37°C) hybridization solution. The chips were hybridized overnight at 44°C. Chips were washed in 1X SSPET and 6X SSPET followed by staining with 2 µg/ml SARPE and 0.5 mg/ml acetylated BSA in 200 µl of 6X SSPET for 8 minutes at room temperature. Chips were scanned using a Molecular Dynamics scanner.

Chip image files were analyzed using Ulysses (Affymetrix, CA) which uses four algorithms to identify potential polymorphisms. Candidate polymorphisms were visually inspected and assigned a confidence value: high confidence candidates displayed all three genotypes, while likely candidates showed only two genotypes (homozygous for reference sequence and heterozygous for reference and variant). Some of the candidate polymorphisms were confirmed by ABI sequencing. Identified polymorphisms were compared to several databases to determine if they were novel. Results are shown in the Table.

Association of Thrombospondin Gene Polymorphisms with Vascular Disease

To determine pivotal genes associated with premature coronary artery disease, we analyzed DNA from 347 patients with MI or coronary revascularization before age 40 (men) or 45 (women) and 422 general population controls. Cases were

drawn (one per family) from a retrospective collection of sibling pairs with premature CAD. Controls were ascertained through random-digit dialing. Both cases and controls were Caucasian. A complete database of phenotypic and laboratory variables for the affected patients afforded logistic regression to control for age, diabetes, body mass index, gender.

Thrombospondin (TSP) 4 and 1 emerged as important SNPs associated with premature CAD and MI. For CAD, 148 of 347 patients carried at least one copy of the TSP-4 variant compared with 142 of 422 control subjects; adjusted odds ratio 1.47, p=0.01. For premature MI, the association was even stronger: 91 of 187 cases vs. 142 of 422 controls had the variant; adjusted odds ratio 2.08, p=0.0003. The TSP-1 SNP was rare. Nonetheless, homozygosity for the variant allele gave an adjusted odds ratio of 9.5, p=.04.

Poly ID	WIAF ID	Genbank or TIGR Accession Number	Position in Sequence	Gene Description	Flanking Seq	Mutation Type	Ref NT	Alt NT	Ref AA	Alt AA
AT3a7	WIAF-13246	U11270	11918	AT3, antithrombin III	CTGCAGGAGT[G/A]GCTGGATGAA	N	G	A	W	*
DRD5u22	WIAF-12913	M67439	310	DRD1, dopamine receptor D1	CATCTGGACC[C/T]TGCTGGGCAA	S	C	T	L	L
DRD5u23	WIAF-12914	M67439	332	DRD1, dopamine receptor D1	GTGCTGGTGT[G/C]CGCAGCCATC	M	G	C	C	S
DRD5u24	WIAF-12915	M67439	369	DRD1, dopamine receptor D1	TGCGCGCCAA[C/G]ATGACCAACG	M	C	G	N	K
DRD5u25	WIAF-12916	M67439	522	DRD1, dopamine receptor D1	TGTGCTCCAC[T/C]GCCTCCATCC	S	T	C	T	T
DRD5u26	WIAF-12917	M67439	953	DRD1, dopamine receptor D1	GCAGAGCACG[C/T]GCAGAGCTGC	M	C	T	A	V
DRD5u27	WIAF-12918	M67439	635	DRD1, dopamine receptor D1	ATGGTCGGCC[T/C]GGCATGGACC	M	T	C	L	P
DRD5u28	WIAF-12919	M67439	606	DRD1, dopamine receptor D1	GCAAGATGAC[T/C]CAGCGCATGG	S	T	C	T	T
DRD5u29	WIAF-12920	M67439	845	DRD1, dopamine receptor D1	TCGCTCATCA[G/A]CTTCTACATC	M	G	A	S	N
DRD5u30	WIAF-12921	M67439	720	DRD1, dopamine receptor D1	GGGGCGGGCT[G/T]GACCTGCCAA	S	G	T	L	L
DRD5u31	WIAF-12922	M67439	1044	DRD1, dopamine receptor D1	AGACCCTGTC[G/A]GTGATCATGG	S	G	A	S	S
DRD5u32	WIAF-12923	M67439	766	DRD1, dopamine receptor D1	GGAGGAGGAC[T/G]TTTGGGAGCC	M	T	G	F	V

DRD5u33	WIAF-12924	M67439	777	DRD1, dopamine receptor D1	TTTGGGAGCC[C/T]GACGTGAATG	S	C	T	P	P
DRD5u34	WIAF-12925	M67439	786	DRD1, dopamine receptor D1	CCGACGTGAA[T/G]GCAGAGAACT	M	T	G	N	K
DRD5u35	WIAF-12926	M67439	887	DRD1, dopamine receptor D1	ACCTACACGC[G/A]CATCTACCGC	M	G	A	R	H
DRD5u36	WIAF-12927	M67439	1279	DRD1, dopamine receptor D1	GTGCAGCCAC[T/G]TCTGCTCCCG	M	T	G	F	V
DRD5u37	WIAF-12928	M67439	1370	DRD1, dopamine receptor D1	GAAATCGCAG[C/T]TGCCTACATC	M	C	T	A	V
DRD5u38	WIAF-12929	M67439	1500	DRD1, dopamine receptor D1	ACCCTGTTGC[T/A]GAGTCTGTCT	S	T	A	A	A
DRD5u39	WIAF-12930	M67439	1338	DRD1, dopamine receptor D1	TCTCCTACAA[C/T]CAAGACATCG	S	C	T	N	N
DRD5u40	WIAF-12931	M67439	1215	DRD1, dopamine receptor D1	CACTCAACCC[C/A]GTCATCTATG	S	C	A	P	P
DRD5u41	WIAF-12932	M67439	1242	DRD1, dopamine receptor D1	ACGCCGACTT[T/C]CAGAAGGTGT	S	T	C	F	F
DRD5u42	WIAF-12933	M67439	1441	DRD1, dopamine receptor D1	CGAGGAGGAG[G/A]GTCCTTTCGA	M	G	A	G	S
DRD5u43	WIAF-12934	M67439	1460	DRD1, dopamine receptor D1	GATCGCATGT[T/C]CCAGATCTAT	M	T	C	F	S
DRD5u44	WIAF-12960	M67439	399	DRD1, dopamine receptor D1	TGTCTCTGGC[C/T]GTGTCTGACC	S	C	T	A	A
DRD5u45	WIAF-12961	M67439	162	DRD1, dopamine receptor D1	TGCCGCCAGG[C/G]AGCAACGGCA	S	C	G	G	G
DRD5u46	WIAF-12962	M67439	195	DRD1, dopamine receptor D1	GGCAGTTCGC[T/G]CTATACCAGC	S	T	G	A	A
DRD5u47	WIAF-12963	M67439	264	DRD1, dopamine receptor D1	TGGGGCCCTC[A/G]CAGGTGGTCA	S	A	G	S	S
DRD5u48	WIAF-12964	M67439	465	DRD1, dopamine receptor D1	TGGCCGGTTA[C/T]TGGCCCTTTG	S	C	T	Y	Y
DRD5u49	WIAF-12965	M67439	511	DRD1, dopamine receptor D1	CTTCGACATC[A/T]TGTGCTCCAC	M	A	T	M	L
DRD5u50	WIAF-12966	M67439	557	DRD1, dopamine receptor D1	ATCAGCGTGG[A/G]CCGCTACTGG	M	A	G	D	G
DRD5u51	WIAF-12967	M67439	476	DRD1, dopamine receptor D1	TGGCCCTTTG[G/A]AGCGTTCTGC	M	G	A	G	E

DRD5u52	WIAF-12968	M67439	1004	DRD1, dopamine receptor D1	AGCCTGCGCG[C/T]TTCCATCAAG	M	C	T	A	V
DRD5u53	WIAF-12969	M67439	1036	DRD1, dopamine receptor D1	GGTTCTCAAG[A/C]CCCTGTCGGT	M	A	C	T	P
DRD5u54	WIAF-12970	M67439	859	DRD1, dopamine receptor D1	CTACATCCCC[G/A]TTGCCATCAT	M	G	A	V	I
DRD5u55	WIAF-12971	M67439	931	DRD1, dopamine receptor D1	GATTTCCTCC[C/T]TGGAGAGGGC	S	C	T	L	L
G10u1	WIAF-10234	J04111	1308	JUN, v-jun avian sarcoma virus 17 oncogene homolog	CCCTCAACGC[C/T]TCGTTCCTCC	S	C	T	A	A
G10u2	WIAF-10235	J04111	1471	JUN, v-jun avian sarcoma virus 17 oncogene homolog	GCTGCTCAAG[C/T]TGGCGTCGCC	S	C	T	L	L
G10u3	WIAF-10253	J04111	2010	JUN, v-jun avian sarcoma virus 17 oncogene homolog	TGGAGTCCCA[G/A]GAGCGGATCA	S	G	A	Q	Q
G1001u1	WIAF-13746	D26135	993	DGKG, diacylglycerol kinase, gamma (90kD)	CCCCAGTGGT[G/A]TACCTGAAGG	S	G	A	V	V
G1001u2	WIAF-13764	D26135	2313	DGKG, diacylglycerol kinase, gamma (90kD)	ATGTGATGAG[A/T]GAGAAACATC	M	A	T	R	S
G1002u1	WIAF-13918	X57206	334	ITPKB, inositol 1,4,5-trisphosphate 3-kinase B	CCCCAAGATC[A/C]GGACAAGCCT	M	A	C	Q	P
G1002u2	WIAF-13925	X57206	575	ITPKB, inositol 1,4,5-trisphosphate 3-kinase B	CCAACTCAGC[T/C]TTCCTGCATA	S	T	C	A	A
G1004u1	WIAF-13567	L36151	1854	PIK4CA, phosphatidylinositol 4-kinase, catalytic, alpha polypeptide	GCCGCTCAGA[C/T]TCCGAGGATG	S	C	T	D	D
G1006u1	WIAF-12375	HT2690	858	PRKCA, protein kinase C, alpha	GGTACAAGTT[G/A]CTTAACCAAG	S	G	A	L	L
G1008u1	WIAF-12397	HT2136	300	PRKCZ, protein kinase C, zeta	CTGGCCTGCC[A/G]TGTCCGGGAG	S	A	G	P	P
G1008u2	WIAF-12398	HT2136	246	PRKCZ, protein kinase C, zeta	AGTGCAGGGA[T/C]GAAGGCCTCA	S	T	C	D	D
G1008u3	WIAF-12399	HT2136	504	PRKCZ, protein kinase C, zeta	GCTGCCACGG[C/T]CTCGTCCCGC	S	C	T	G	G
G1008u4	WIAF-12403	HT2136	807	PRKCZ, protein kinase C, zeta	AGAAGAATGA[C/T]CAAATTTACG	S	C	T	D	D
G1008u5	WIAF-12404	HT2136	1514	PRKCZ, protein kinase C, zeta	GGATTTTCTG[A/T]CATCAAGTCC	M	A	T	D	V

G1008u6	WIAF-12412	HT2136	166	PRKCZ, protein kinase C, zeta	CAAGTGGGTG[G/A]ACAGCGAAGG	M	G	A	D	N
G1008u7	WIAF-12418	HT2136	560	PRKCZ, protein kinase C, zeta	TCCCAAGAGC[C/T]TCCAGTAGAC	M	C	T	P	L
G1009u1	WIAF-12396	L05186	2495	PTK2, PTK2 protein tyrosine kinase 2	TCATCAACAA[G/A]ATGAAACTGG	S	G	A	K	K
G1011u1	WIAF-11988	X07876	1250	WNT2, wingless-type MMTV integration site family member 2	TCCCATGTCA[C/A]CCGGATGACC	M	C	A	T	N
G1011u2	WIAF-11997	X07876	788	WNT2, wingless-type MMTV integration site family member 2	GACTATGGGA[T/C]CAAATTTGCC	M	T	C	I	T
G1011u3	WIAF-12014	X07876	1338	WNT2, wingless-type MMTV integration site family member 2	TGCACACATG[C/A]AAGGCCCCCA	N	C	A	C	*
G1011u4	WIAF-13475	X07876	856	WNT2, wingless-type MMTV integration site family member 2	CCTGATGAAT[C/T]TTCACAACAA	M	C	T	L	F
G1011u5	WIAF-13476	X07876	958	WNT2, wingless-type MMTV integration site family member 2	GACATGCTGG[C/T]TGGCCATGGC	S	C	T	L	L
G1011u6	WIAF-13477	X07876	789	WNT2, wingless-type MMTV integration site family member 2	ACTATGGGAT[C/T]AAATTTGCCC	S	C	T	I	I
G1011u7	WIAF-13478	X07876	823	WNT2, wingless-type MMTV integration site family member 2	TGCAAAGGAA[A/G]GGAAAGGAAA	M	A	G	R	G
G1012u1	WIAF-12408	HT48910	1574	WNT2B, wingless-type MMTV integration site family, member 2B	ATACTTGCAA[A/G]GCCCCCAAGA	S	A	G	K	K
G1016a1	WIAF-12125	Z22534	793	ACVR1, activin A receptor, type I	GGCAAGGGGA[A/G]AATGTTGCCG	S	A	G	E	E
G1016u2	WIAF-12392	Z22534	373	ACVR1, activin A receptor, type I	CTGGCCAAGC[T/C]GTGGAGTGCT	S	T	C	A	A
G1018u1	WIAF-12413	X74210	1150	ADCY2, adenylate cyclase 2 (brain)	CAAATTGCGA[G/T]TGGGTATTAA	M	G	T	V	L
G1019u1	WIAF-12394	U83867	5475	SPTAN1, spectrin, alpha, non-erythrocytic 1 (alpha-fodrin)	GGGACCTAAC[T/C]GGCGTGCAGA	S	T	C	T	T

G1019u2	WIAF-12406	U83867	1223	SPTAN1, spectrin, alpha, non-erythrocytic 1 (alpha-fodrin)	GCCCTCATCA[A/G]TGCAGATGAG	M	A	G	N	S
G1019u3	WIAF-12409	U83867	3555	SPTAN1, spectrin, alpha, non-erythrocytic 1 (alpha-fodrin)	CTGAAGGTCT[T/C]ATGGCAGAGG	S	T	C	L	L
G1019u4	WIAF-12415	U83867	3369	SPTAN1, spectrin, alpha, non-erythrocytic 1 (alpha-fodrin)	TCCGTGAAGC[G/A]AATGAACTAC	S	G	A	A	A
G1019u5	WIAF-12417	U83867	5839	SPTAN1, spectrin, alpha, non-erythrocytic 1 (alpha-fodrin)	TGAGACAGAC[T/A]TCACCGTCCA	M	T	A	F	I
G1022u1	WIAF-12393	U45945	631	ATP1B2, ATPase, Na+/K+ transporting, beta 2 polypeptide	CATGAATGTT[A/G]CCTGTGCTGG	M	A	G	T	A
G1022u2	WIAF-12400	U45945	432	ATP1B2, ATPase, Na+/K+ transporting, beta 2 polypeptide	GCCGCCCTGG[G/A]CGCTATTACG	S	G	A	G	G
G1023u1	WIAF-12401	D89722	395	ARNTL, aryl hydrocarbon receptor nuclear translocator-like	AACATTAAGA[G/C]GTGCCACCAA	M	G	C	G	R
G1023u2	WIAF-12407	D89722	681	ARNTL, aryl hydrocarbon receptor nuclear translocator-like	CTCATAGATG[C/T]AAAAACTGGA	M	C	T	A	V
G1024u1	WIAF-12410	U85946	731	Homo sapiens brain secretory protein hSec10p (HSEC10) mRNA, complete cds.	GATAGATTTT[C/T]AGAAGTTAAA	M	C	T	S	L
G1027u1	WIAF-12402	L47647	1135	CKB, creatine kinase, brain	TCGAGATGGA[A/G]CAGCGGCTGG	S	A	G	E	E
G1027u2	WIAF-12405	L47647	499	CKB, creatine kinase, brain	GGGAGCGCCG[A/C]GCCATCGAGA	S	A	C	R	R
G103u1	WIAF-10427	HT2269	335	ERCC5, excision repair cross-complementing rodent repair deficiency, complementation group 5 (xeroderma pigmentosum, complementation group G (Cockayne syndrome))	GGGATCGCCA[T/C]GGGAACTCAA	S	T	C	H	H

G103u2	WIAF-10429	HT2269	1221	ERCC5, excision repair cross-complementing rodent repair deficiency, complementation group 5 (xeroderma pigmentosum, complementation group G (Cockayne syndrome))	CCCTCCTTCT[C/T]CAAGAACTTT	M	C	T	C	P	S										
G103u3	WIAF-10431	HT2269	1783	ERCC5, excision repair cross-complementing rodent repair deficiency, complementation group 5 (xeroderma pigmentosum, complementation group G (Cockayne syndrome))	TCTCCAACTT[G/C]TACAAATTCT	M	G	C	C	C	S										
G103u4	WIAF-10432	HT2269	2077	ERCC5, excision repair cross-complementing rodent repair deficiency, complementation group 5 (xeroderma pigmentosum, complementation group G (Cockayne syndrome))	ACTGAATCTG[C/A]AGGCCAGGAT	M	C	A	A	A	E										
G103u5	WIAF-10446	HT2269	3338	ERCC5, excision repair cross-complementing rodent repair deficiency, complementation group 5 (xeroderma pigmentosum, complementation group G (Cockayne syndrome))	AATTTGAGCT[A/T]CTTGATAAGG	S	A	T	L	L	L										
G103u6	WIAF-10447	HT2269	3487	ERCC5, excision repair cross-complementing rodent repair deficiency, complementation group 5 (xeroderma pigmentosum, complementation group G (Cockayne syndrome))	TCAGAATCAT[C/T]TGATGGATCT	M	C	T	S	F											

G103u7	WIAF-10448	HT2269	3507	ERCC5, excision repair cross-complementing rodent repair deficiency, complementation group 5 (xeroderma pigmentosum, complementation group G (Cockayne syndrome))	TTCAAGTGAA[C/G]ATGCTGAAAG	M	C	G	H	D
G103u8	WIAF-10457	HT2269	1388	ERCC5, excision repair cross-complementing rodent repair deficiency, complementation group 5 (xeroderma pigmentosum, complementation group G (Cockayne syndrome))	CTCTTGACGA[T/G]GACGAAGATG	M	T	G	D	E
G103u9	WIAF-10458	HT2269	1362	ERCC5, excision repair cross-complementing rodent repair deficiency, complementation group 5 (xeroderma pigmentosum, complementation group G (Cockayne syndrome))	CCGGACTCTT[T/C]CAGCCATTAA	M	T	C	S	P
G103u10	WIAF-10459	HT2269	2357	ERCC5, excision repair cross-complementing rodent repair deficiency, complementation group 5 (xeroderma pigmentosum, complementation group G (Cockayne syndrome))	CTGAGAAAGA[T/C]GCGGAAGATT	S	T	C	D	D
G103u11	WIAF-10462	HT2269	3109	ERCC5, excision repair cross-complementing rodent repair deficiency, complementation group 5 (xeroderma pigmentosum, complementation group G (Cockayne syndrome))	TGGAACAGAA[C/T]GAAGACAGAT	M	C	T	T	M

G103u12	WIAF-10463	HT2269	3138	ERCC5, excision repair cross-complementing rodent repair deficiency, complementation group 5 (xeroderma pigmentosum, complementation group G (Cockayne syndrome))	GTTTCCTGTA[T/C]TAAAGCAACT	S	T	C	L	L
G103u14	WIAF-10484	HT2269	3553	ERCC5, excision repair cross-complementing rodent repair deficiency, complementation group 5 (xeroderma pigmentosum, complementation group G (Cockayne syndrome))	AGAACAGCTG[C/T]GAAAGAGCCA	M	C	T	A	V
G103u15	WIAF-10485	HT2269	1429	ERCC5, excision repair cross-complementing rodent repair deficiency, complementation group 5 (xeroderma pigmentosum, complementation group G (Cockayne syndrome))	GATGTGCAGA[C/T]GGGAGGGCCA	M	C	T	T	M
G103a16	WIAF-12097	HT2269	3335	ERCC5, excision repair cross-complementing rodent repair deficiency, complementation group 5 (xeroderma pigmentosum, complementation group G (Cockayne syndrome))	AAGAATTTGA[G/T]CTACTTGATA	M	G	T	E	D
G1030u1	WIAF-12411	U07358	203	ZPK, zipper (leucine) protein kinase	ACACTTCTGA[C/T]TGCACTCCCG	S	C	T	D	D
G1030u2	WIAF-12416	U07358	1806	ZPK, zipper (leucine) protein kinase	GCCACCCCAT[G/T]AACCTGGAGG	N	G	T	E	*
G1031a1	WIAF-12124	U87460	2825	GPR37, G protein-coupled receptor 37 (endothelin receptor type B-like)	GAGTCACCAC[C/T]TTCACCTTAT	S	C	T	T	T
G1032u1	WIAF-12381	U57911	926	C11ORF8, chromosome 11 open reading frame 8	ACGTACATCA[A/C]TGCCTCGACG	M	A	C	N	T

G1033u1	WIAF-12437	M65188	431	GJA1, gap junction protein, alpha 1, 43kD (connexin 43)	TCTGTACCCA[C/T]ACTCTTGTAC	M	C	T	T	I
G1033u2	WIAF-12438	M65188	169	GJA1, gap junction protein, alpha 1, 43kD (connexin 43)	AGGCAACATG[G/C]GTGACTGGAG	M	G	C	G	R
G1033u3	WIAF-12439	M65188	467	GJA1, gap junction protein, alpha 1, 43kD (connexin 43)	TATGTGATGC[G/A]AAAGGAAGAG	M	G	A	R	Q
G1033u4	WIAF-12440	M65188	263	GJA1, gap junction protein, alpha 1, 43kD (connexin 43)	TTCATTTTCC[G/A]AATCCTGCTG	M	G	A	R	Q
G1033u5	WIAF-12441	M65188	218	GJA1, gap junction protein, alpha 1, 43kD (connexin 43)	CAAGCCTACT[C/T]AACTGCTGGA	M	C	T	S	L
G1033u6	WIAF-12442	M65188	498	GJA1, gap junction protein, alpha 1, 43kD (connexin 43)	AGAAAGAGGA[A/G]GAACTCAAGG	S	A	G	E	E
G1033u7	WIAF-12465	M65188	550	GJA1, gap junction protein, alpha 1, 43kD (connexin 43)	GCACTTGAAG[C/A]AGATTGAGAT	M	C	A	Q	K
G1033u8	WIAF-12466	M65188	548	GJA1, gap junction protein, alpha 1, 43kD (connexin 43)	ATGCACTTGA[A/G]GCAGATTGAG	M	A	G	K	R
G1033u9	WIAF-12486	M65188	933	GJA1, gap junction protein, alpha 1, 43kD (connexin 43)	CGCTGAGCCC[T/C]GCCAAAGACT	S	T	C	P	P
G1033u10	WIAF-12487	M65188	990	GJA1, gap junction protein, alpha 1, 43kD (connexin 43)	CCTCACCAAC[C/T]GCTCCCCTCT	S	C	T	T	T
G1033u11	WIAF-12488	M65188	1034	GJA1, gap junction protein, alpha 1, 43kD (connexin 43)	AAGCTGGTTA[C/A]TGGCGACAGA	M	C	A	T	N
G1033u12	WIAF-12489	M65188	1158	GJA1, gap junction protein, alpha 1, 43kD (connexin 43)	CTAACTCCCA[T/C]GCACAGCCTT	S	T	C	H	H
G1033u13	WIAF-12490	M65188	1222	GJA1, gap junction protein, alpha 1, 43kD (connexin 43)	TGGACATGAA[T/C]TACAGCCACT	S	T	C	L	L

G1033u14	WIAF-12491	M65188	1069	GJA1, gap junction protein, alpha 1, 43kD (connexin 43)	CCGCAATTAC[A/G]ACAAGCAAGC	M	A	G	N	D
G1033u15	WIAF-12492	M65188	1250	GJA1, gap junction protein, alpha 1, 43kD (connexin 43)	GTGGACCAGC[G/A]ACCTTCAAGC	M	G	A	R	Q
G1033u16	WIAF-12496	M65188	423	GJA1, gap junction protein, alpha 1, 43kD (connexin 43)	TATTTGTGTC[T/C]GTACCCACAC	S	T	C	S	S
G1033u17	WIAF-12503	M65188	880	GJA1, gap junction protein, alpha 1, 43kD (connexin 43)	CGTTAAGGAT[C/T]GGGTTAAGGG	M	C	T	R	W
G1033u18	WIAF-12504	M65188	855	GJA1, gap junction protein, alpha 1, 43kD (connexin 43)	AACTCTTCTA[T/C]GTTTTCTTCA	S	T	C	Y	Y
G1033u19	WIAF-12505	M65188	576	GJA1, gap junction protein, alpha 1, 43kD (connexin 43)	AGTTCAAGTA[C/T]GGTATTGAAG	S	C	T	Y	Y
G1033u20	WIAF-12512	M65188	1255	GJA1, gap junction protein, alpha 1, 43kD (connexin 43)	CCAGCGACCT[T/G]CAAGCAGAGC	M	T	G	S	A
G1033u21	WIAF-12513	M65188	1078	GJA1, gap junction protein, alpha 1, 43kD (connexin 43)	CAACAAGCAA[G/A]CAAGTGAGCA	M	G	A	A	T
G1033u22	WIAF-12514	M65188	1097	GJA1, gap junction protein, alpha 1, 43kD (connexin 43)	CAAAACTGGG[C/G]TAATTACAGT	M	C	G	A	G
G1034u1	WIAF-12443	J03544	1201	PYGB, phosphorylase, glycogen; brain	AGACCTGTGC[A/G]TACACCAACC	S	A	G	A	A
G1034u2	WIAF-12469	J03544	771	PYGB, phosphorylase, glycogen; brain	GACACCCCAG[T/C]GCCCGGCTAC	M	T	C	V	A
G1034u3	WIAF-12470	J03544	1465	PYGB, phosphorylase, glycogen; brain	TCCACTCGGA[G/C]ATCGTGAAAC	M	G	C	E	D
G1034u4	WIAF-12471	J03544	1583	PYGB, phosphorylase, glycogen; brain	GGGGCTGGCC[G/A]ATACCATCGT	M	G	A	D	N
G1034u5	WIAF-12472	J03544	1774	PYGB, phosphorylase, glycogen; brain	CCATGTTCGA[T/C]GTGCATGTGA	S	T	C	D	D
G1034u6	WIAF-12474	J03544	2449	PYGB, phosphorylase, glycogen; brain	AGGTGGACCA[G/A]CTGTACCGGA	S	G	A	Q	Q

G1034u7	WIAF-12508	J03544	718	PYGB, phosphorylase, glycogen; brain	CCCCCGACGG[C/T]GTGAAGTGGC	S	C	T	G	G
G1035u1	WIAF-12484	U97105	1962	DPYSL2, dihydropyrimidinase-like 2	GCAGAGGAGC[A/G]GCAGAGGATC	M	A	G	Q	R
G1035u2	WIAF-12485	U97105	2842	DPYSL2, dihydropyrimidinase-like 2	ATGACGGACC[T/C]GTGTGTGAAG	S	T	C	P	P
G1035u3	WIAF-12511	U97105	2062	DPYSL2, dihydropyrimidinase-like 2	CCATCACCAT[C/T]GCCAACCAGA	S	C	T	I	I
G1036u1	WIAF-12444	D88460	311	WASL, Wiskott-Aldrich syndrome-like	ACGTGGGGTC[C/T]CTGTTGCTCA	S	C	T	S	S
G1038u1	WIAF-12445	HT2746	994	PCTK2, PCTAIRE protein kinase 2	TAGAAGAAAG[G/A]TATTGCATCG	M	G	A	V	I
G1039u1	WIAF-12429	HT2747	955	serine/threonine kinase, PCTAIRE-3	ATCCAAGAGT[C/T]GCATGTCAGC	M	C	T	R	C
G1039u2	WIAF-12458	HT2747	808	serine/threonine kinase, PCTAIRE-3	CACAGAAGAG[A/T]CGTGGCCCGG	M	A	T	T	S
G1041u1	WIAF-12459	X72886	544	H.sapiens TYRO3 mRNA.	CAAGTGGCTG[G/C]CCCTGGAGAG	M	G	C	A	P
G1041u2	WIAF-12460	X72886	693	H.sapiens TYRO3 mRNA.	TTGGCGGGAA[C/T]CGCCTGAAAC	S	C	T	N	N
G1041u3	WIAF-12502	X72886	561	H.sapiens TYRO3 mRNA.	AGAGCCTGGC[C/T]GACAACCTGT	S	C	T	A	A
G1043u1	WIAF-12448	M94055	5481	Human voltage-gated sodium channel mRNA, complete cds.	CTCTGAGTGA[G/A]GATGACTTTG	S	G	A	E	E
G1043u2	WIAF-12449	M94055	5205	Human voltage-gated sodium channel mRNA, complete cds.	TTGAGACCTT[T/C]GGCAACAGCA	S	T	C	F	F
G1043u3	WIAF-12450	M94055	5224	Human voltage-gated sodium channel mRNA, complete cds.	CATGATCTGC[C/T]TGTTCCAAAT	S	C	T	L	L
G1043u4	WIAF-12451	M94055	5514	Human voltage-gated sodium channel mRNA, complete cds.	AGGTTTGGGA[G/A]AAGTTTGATC	S	G	A	E	E
G1043u5	WIAF-12452	M94055	5217	Human voltage-gated sodium channel mRNA, complete cds.	GCAACAGCAT[G/C]ATCTGCCTGT	M	G	C	M	I
G1043u6	WIAF-12453	M94055	5334	Human voltage-gated sodium channel mRNA, complete cds.	GCTCAGTTAA[A/G]GGAGACTGTG	S	A	G	K	K

G1043u7	WIAF-12454	M94055	5424	Human voltage-gated sodium channel mRNA, complete cds.	TGTACATCGC[G/C]GTCATCCTGG	S	G	C	A	A
G1043u8	WIAF-12455	M94055	5322	Human voltage-gated sodium channel mRNA, complete cds.	ATCACCCTGG[A/C]AGCTCAGTTA	S	A	C	G	G
G1043u9	WIAF-12456	M94055	1200	Human voltage-gated sodium channel mRNA, complete cds.	ATGGCTACAC[G/A]AGCTTTGACA	S	G	A	T	T
G1043u10	WIAF-12499	M94055	1170	Human voltage-gated sodium channel mRNA, complete cds.	TCTGTGTGAA[G/T]GCTGGTAGAA	M	G	T	K	N
G1046a1	WIAF-13187	U50352	267	ACCN1, amiloride-sensitive cation channel 1, neuronal (degenerin)	TCCCAGCTGT[G/A]ACCCTCTGTA	S	G	A	V	V
G1046a2	WIAF-13188	U50352	282	ACCN1, amiloride-sensitive cation channel 1, neuronal (degenerin)	TCTGTAACCT[C/g]AATGGCTTCC	S	C	g	L	L
G1046a3	WIAF-13189	U50352	315	ACCN1, amiloride-sensitive cation channel 1, neuronal (degenerin)	TCACCACCAA[C/t]GACCTGTACC	S	C	t	N	N
G1046a4	WIAF-13190	U50352	386	ACCN1, amiloride-sensitive cation channel 1, neuronal (degenerin)	CCCCATCTGG[C/a]TGACCCCTCC	M	C	a	A	D
G1046a5	WIAF-13191	U50352	417	ACCN1, amiloride-sensitive cation channel 1, neuronal (degenerin)	CCCTGCGGCA[G/A]AAGGCCAACT	S	G	A	Q	Q
G1048u1	WIAF-12641	HT5174S	3214	REST, RE1-silencing transcription factor	CAGTCAAAGC[G/A]GCTAAGGGAG	S	G	A	A	A
G1048u2	WIAF-12642	HT5174S	3199	REST, RE1-silencing transcription factor	CAAAGGAAGC[C/G]TTGGCAGTCA	S	C	G	A	A
G1048u3	WIAF-12657	HT5174S	2125	REST, RE1-silencing transcription factor	CTCCCATGGA[G/T]ACTGCTCAGA	M	G	T	E	D
G1048u4	WIAF-12660	HT5174S	2333	REST, RE1-silencing transcription factor	GGAACCTGTT[A/C]AGATAGAGCT	M	A	C	K	Q
G1051u1	WIAF-12431	HT28321	658	SCNN1G, sodium channel, nonvoltage-gated 1, gamma	ATGACACCTC[C/T]GACTGTGCCA	S	C	T	S	S
G1051u2	WIAF-12434	HT28321	1735	SCNN1G, sodium channel, nonvoltage-gated 1, gamma	AAGCCAAGGA[G/A]TGGTGGGCCT	S	G	A	E	E

G1051u3	WIAF-12473	HT28321	409	SCNN1G, sodium channel, nonvoltage-gated 1, gamma	AGTCCCTGTA[T/C]GGCTTTCCAG	S	T	C	Y	Y
G1051u4	WIAF-12475	HT28321	953	SCNN1G, sodium channel, nonvoltage-gated 1, gamma	AGTCATTTTG[T/C]ACATAAACGA	M	T	C	Y	H
G1051u5	WIAF-12476	HT28321	975	SCNN1G, sodium channel, nonvoltage-gated 1, gamma	GAGGAATACA[A/G]CCCATTCCTC	M	A	G	N	S
G1051u6	WIAF-12477	HT28321	1192	SCNN1G, sodium channel, nonvoltage-gated 1, gamma	CTGCCTACTC[G/A]CTCCAGATCT	S	G	A	S	S
G1053a1	WIAF-13192	HT2201	4085	SCN5A, sodium channel, voltage-gated, type V, alpha polypeptide (long (electrocardiographic) QT syndrome 3)	CGTCCTCTGA[G/A]AGCTCTGTCA	M	G	A	R	K
G1053a2	WIAF-13193	HT2201	5607	SCN5A, sodium channel, voltage-gated, type V, alpha polypeptide (long (electrocardiographic) QT syndrome 3)	ACTTTGCCGA[C/T]GCCCTGTCTG	S	C	T	D	D
G1053a3	WIAF-13194	HT2201	5828	SCN5A, sodium channel, voltage-gated, type V, alpha polypeptide (long (electrocardiographic) QT syndrome 3)	GAGCCCATCA[C/T]CACCACACTC	M	C	T	T	I
G1053a4	WIAF-13202	HT2201	713	SCN5A, sodium channel, voltage-gated, type V, alpha polypeptide (long (electrocardiographic) QT syndrome 3)	GCGTTCACTT[T/A]CCTTCGGGAC	M	T	A	F	Y
G1053a5	WIAF-13203	HT2201	6148	SCN5A, sodium channel, voltage-gated, type V, alpha polypeptide (long (electrocardiographic) QT syndrome 3)	CCACAGTGAA[G/T]ATCTCGCCGA	M	G	T	D	Y
G1053a6	WIAF-13204	HT2201	6217	SCN5A, sodium channel, voltage-gated, type V, alpha polypeptide (long (electrocardiographic) QT syndrome 3)	GGCCTGGCTG[G/T]CCAGGACACA	-	G	T	-	-

G1053a7	WIAF-13205	HT2201	6324	SCN5A, sodium channel, voltage-gated, type V, alpha polypeptide (long (electrocardiographic) QT syndrome 3)	AATGGGCCTC[G/A]GCCCCGCGGA	-	G	A	-	-
G1054u1	WIAF-12419	HT2202	2252	SCN4A, sodium channel, voltage-gated, type IV, alpha polypeptide	TTGGCAAGAG[C/T]TACAAGGAGT	S	C	T	S	S
G1054u2	WIAF-12423	HT2202	4559	SCN4A, sodium channel, voltage-gated, type IV, alpha polypeptide	TGGTCATGTT[C/T]ATCTACTCCA	S	C	T	F	F
G1054u3	WIAF-12424	HT2202	4856	SCN4A, sodium channel, voltage-gated, type IV, alpha polypeptide	TCAACATGTA[C/G]ATCGCCATCA	N	C	G	Y	*
G1054u4	WIAF-12425	HT2202	4777	SCN4A, sodium channel, voltage-gated, type IV, alpha polypeptide	GTCAAGGGTG[A/G]CTGCGGCAAC	M	A	G	D	G
G1054u5	WIAF-12426	HT2202	4863	SCN4A, sodium channel, voltage-gated, type IV, alpha polypeptide	GTACATCGCC[A/G]TCATCCTGGA	M	A	G	I	V
G1054u6	WIAF-12427	HT2202	4566	SCN4A, sodium channel, voltage-gated, type IV, alpha polypeptide	GTTCATCTAC[T/G]CCATCTTCGG	M	T	G	S	A
G1054u7	WIAF-12428	HT2202	4923	SCN4A, sodium channel, voltage-gated, type IV, alpha polypeptide	TGGTGAAGAT[G/T]ACTTTGAGAT	M	G	T	D	Y
G1054u8	WIAF-12446	HT2202	3595	SCN4A, sodium channel, voltage-gated, type IV, alpha polypeptide	TTCTGGCTGA[T/C]CTTCAGCATC	M	T	C	I	T
G1054u9	WIAF-12447	HT2202	4203	SCN4A, sodium channel, voltage-gated, type IV, alpha polypeptide	GGAGACAGAC[G/A]ACCAGAGCCA	M	G	A	D	N
G1054u10	WIAF-12495	HT2202	4811	SCN4A, sodium channel, voltage-gated, type IV, alpha polypeptide	TCTGCTTCTT[C/A]TGCAGCTATA	M	C	A	F	L
G1054u11	WIAF-12497	HT2202	5555	SCN4A, sodium channel, voltage-gated, type IV, alpha polypeptide	CAGGGCAGAC[T/G]GTGCGCCCAG	S	T	G	T	T

G1054u12	WIAF-12498	HT2202	5480	SCN4A, sodium channel, voltage-gated, type IV, alpha polypeptide	CAGGGGACGC[C/T]GGACCCACTA	S	C	T	A	A
G1059u1	WIAF-12432	HT33704	112	APLP1, amyloid beta (A4) precursor-like protein 1	CGCTGCTGCT[G/A]CCACTATTGC	S	G	A	L	L
G1059u2	WIAF-12433	HT33704	140	APLP1, amyloid beta (A4) precursor-like protein 1	TCTGCGCGCG[C/T]AGCCCGCCAT	N	C	T	Q	*
G1059u3	WIAF-12435	HT33704	1344	APLP1, amyloid beta (A4) precursor-like protein 1	CAGCATGTGG[C/T]CGCCGTGGAT	M	C	T	A	V
G1059u4	WIAF-12457	HT33704	1687	APLP1, amyloid beta (A4) precursor-like protein 1	ATGAGCGAAA[G/A]GTGAATGCGT	S	G	A	K	K
G1059u5	WIAF-12500	HT33704	976	APLP1, amyloid beta (A4) precursor-like protein 1	GGTTCCTGAG[A/G]GCCAAGATGG	S	A	G	R	R
G1059u6	WIAF-12501	HT33704	1786	APLP1, amyloid beta (A4) precursor-like protein 1	GTGAGGCTGT[A/G]TCGGGTCTGC	S	A	G	V	V
G1060u1	WIAF-12436	HT1418	1744	APLP2, amyloid beta (A4) precursor-like protein 2	CCAAGAAATT[C/G]AAGAGGAAAT	M	C	G	Q	E
G1060u2	WIAF-12467	HT1418	2213	APLP2, amyloid beta (A4) precursor-like protein 2	ATCAGCCTGG[T/G]GATGCTGAGG	M	T	G	V	G
G1060u3	WIAF-12468	HT1418	2256	APLP2, amyloid beta (A4) precursor-like protein 2	GCCACGGGAT[C/T]GTGGAGGTTG	S	C	T	I	I
G1066a1	WIAF-13195	HT3538	566	CCKBR, cholecystokinin B receptor	CTTTGGCACC[G/A]TCATCTGCAA	M	G	A	V	I
G1066a2	WIAF-13196	HT3538	607	CCKBR, cholecystokinin B receptor	GGGTGTCTGT[G/A]AGTGTGTCCA	S	G	A	V	V
G1066a3	WIAF-13206	HT3538	864	CCKBR, cholecystokinin B receptor	CTGCTGCTTC[T/A]GCTCTTGTTC	M	T	A	L	Q
G1067u1	WIAF-12478	HT0830	684	KCNA1, potassium voltage-gated channel, shaker-related subfamily, member 1 (episodic ataxia with myokymia)	AAACGCTGTG[C/T]ATCATCTGGT	S	C	T	C	C
G1067u2	WIAF-12479	HT0830	722	KCNA1, potassium voltage-gated channel, shaker-related subfamily, member 1 (episodic ataxia with myokymia)	GTGCGCTTCT[T/C]CGCCTGCCCC	M	T	C	F	S

G1067u3	WIAF-12480	HT0830	804	KCNA1, potassium voltage-gated channel, shaker-related subfamily, member 1 (episodic ataxia with myokymia)	ATTTCATCAC[C/G]CTGGGCACCG	S	C	G	T	T
G1067u4	WIAF-12509	HT0830	690	KCNA1, potassium voltage-gated channel, shaker-related subfamily, member 1 (episodic ataxia with myokymia)	TGTGCATCAT[C/T]TGGTTCTCCT	S	C	T	I	I
G1068u1	WIAF-12493	HT0831	774	KCNA2, potassium voltage-gated channel, shaker-related subfamily, member 2	TGAACATCAT[T/A]GACATTGTGG	S	T	A	I	I
G1070a1	WIAF-13197	HT27728	522	KCNJ6, potassium inwardly-rectifying channel, subfamily J, member 6	CACAGTGACC[T/C]GGCTCTTTTT	M	T	C	W	R
G1070a2	WIAF-13201	HT27728	1244	KCNJ6, potassium inwardly-rectifying channel, subfamily J, member 6	CCCTGGAGGA[T/C]GGGTTCTACG	S	T	C	D	D
G1070a3	WIAF-13207	HT27728	707	KCNJ6, potassium inwardly-rectifying channel, subfamily J, member 6	ATAAATGCCC[G/A]GAGGGAATTA	S	G	A	P	P
G1071u1	WIAF-12422	HT48672	1534	KCNJ3, potassium inwardly-rectifying channel, subfamily J, member 3	TTCCGGGCAA[C/T]TCAGAAGAAA	S	C	T	N	N
G1073u1	WIAF-12461	HT4556	1127	KCNJ1, potassium inwardly-rectifying channel, subfamily J, member 1	CACTGTGCCA[T/C]GTGCCTTTAT	M	T	C	M	T
G1074u1	WIAF-12462	HT27804	289	KCNAB2, potassium voltage-gated channel, shaker-related subfamily, beta member 2	ACCTCTTCGA[T/C]ACAGCAGAAG	S	T	C	D	D
G1079u1	WIAF-12463	HT27383	1130	potassium channel, inwardly rectifing (GB:D50582)	ACCTGGCCGA[T/A]GAGATCCTGT	M	T	A	D	E
G1079u2	WIAF-12464	HT27383	1192	potassium channel, inwardly rectifing (GB:D50582)	CGTTACTCTG[T/G]GGACTACTCC	M	T	G	V	G

G1079u3	WIAF-12481	HT27383	708	potassium channel, inwardly rectifing (GB:D50582)	GCTTGGCTGC[A/G]TCTTCATGAA	M	A	G	I	V
G1079u4	WIAF-12482	HT27383	779	potassium channel, inwardly rectifing (GB:D50582)	CGGTGATCGC[T/C]CTGCGCCACG	S	T	C	A	A
G1079u5	WIAF-12483	HT27383	276	potassium channel, inwardly rectifing (GB:D50582)	GGACCCTGCC[G/A]AGCCCAGGTA	M	G	A	E	K
G1079u6	WIAF-12510	HT27383	489	potassium channel, inwardly rectifing (GB:D50582)	GTGGCTCATC[G/A]CCTTCGCCCA	M	G	A	A	T
G1080u1	WIAF-12536	HT4412	1099	KCNJ4, potassium inwardly-rectifying channel, subfamily J, member 4	TGGACTACTC[A/G]CGTTTTCACA	S	A	G	S	S
G1080u2	WIAF-12537	HT4412	1050	KCNJ4, potassium inwardly-rectifying channel, subfamily J, member 4	GGCCACCGCT[T/A]TGAGCCTGTG	M	T	A	F	Y
G1081u1	WIAF-12538	HT27724	1090	KCNJ2, potassium inwardly-rectifying channel, subfamily J, member 2	GGCCACCGCT[A/T]TGAGCCTGTG	M	A	T	Y	F
G1082u1	WIAF-12662	HT28319	768	potassium channel, inwardly rectifying, high conductance, alpha subunit	CGCGGGTCAC[C/T]GAGGAGGGCG	S	C	T	T	T
G1082u2	WIAF-12663	HT28319	854	potassium channel, inwardly rectifying, high conductance, alpha subunit	CTGGTGTCGC[C/T]CATCACCATC	M	C	T	P	L
G1082u3	WIAF-12679	HT28319	471	potassium channel, inwardly rectifying, high conductance, alpha subunit	TCTCCATCGA[G/C]ACGCAGACCA	M	G	C	E	D
G1084a1	WIAF-13198	HT0383	2028	KCNB1, potassium voltage-gated channel, Shab-related subfamily, member 1	CACTCCCCAG[C/A]AAGACTGGGG	M	C	A	S	R
G1084a2	WIAF-13199	HT0383	2033	KCNB1, potassium voltage-gated channel, Shab-related subfamily, member 1	CCCAGCAAGA[C/G]TGGGGGCAGC	M	C	G	T	S

G1084a3	WIAF-13200	HT0383	2321	KCNB1, potassium voltage-gated channel, Shab-related subfamily, member 1	GAGTGTGCCA [C/A] GCTTTTGGAC	M	C	A	T	K				
G1084a4	WIAF-13208	HT0383	870	KCNB1, potassium voltage-gated channel, Shab-related subfamily, member 1	ACAACCCCCA [G/A] CTGGCCCACG	S	G	A	Q	Q				
G1088u1	WIAF-12516	HT0522	1503	KCNA5, potassium voltage-gated channel, shaker-related subfamily, member 5	TCCTGGGCAA [G/A] ACCTTGCAGG	S	G	A	K	K				
G1088u2	WIAF-12519	HT0522	1249	KCNA5, potassium voltage-gated channel, shaker-related subfamily, member 5	CGAGCTGCTC [G/A] TGCGCTTCTT	M	G	A	V	M				
G1088u3	WIAF-12520	HT0522	973	KCNA5, potassium voltage-gated channel, shaker-related subfamily, member 5	CTCTGGGTCC [G/A] CGCGGGCCAT	M	G	A	A	T				
G1088u4	WIAF-12521	HT0522	1013	KCNA5, potassium voltage-gated channel, shaker-related subfamily, member 5	GTTATCCTCA [T/C] CTCCATCATC	M	T	C	I	T				
G1090u1	WIAF-12651	HT1497	1836	KCNA6, potassium voltage-gated channel, shaker-related subfamily, member 6	CAACCAGCCA [G/A] TGGAGGAGGC	M	G	A	S	N				
G1091u1	WIAF-12714	HT0222	843	KCNA3, potassium voltage-gated channel, shaker-related subfamily, member 3	CATCATCTGG [T/C] TCTCCTTCGA	M	T	C	F	L				
G1094a1	WIAF-13218	HT27381	1280	KCNJ8, potassium inwardly-rectifying channel, subfamily J, member 8	GTGTATTCTG [T/a] GGATTACTCC	M	T	a	V	E				

G1095u1	WIAF-12532	HT2629	765	KCNMA1, potassium large conductance calcium-activated channel, subfamily M, alpha member 1	TTCTCTACTT[C/T]GGCTTGCGGT	S	C	T	F	F
G1095u2	WIAF-12533	HT2629	2441	KCNMA1, potassium large conductance calcium-activated channel, subfamily M, alpha member 1	GTGGTCTGCA[T/C]CTTTGGCGAC	M	T	C	I	T
G1095u3	WIAF-12534	HT2629	2714	KCNMA1, potassium large conductance calcium-activated channel, subfamily M, alpha member 1	GATGATACTT[C/G]GCTGCAGGAC	M	C	G	S	W
G1095u4	WIAF-12535	HT2629	2439	KCNMA1, potassium large conductance calcium-activated channel, subfamily M, alpha member 1	TCGTGGTCTG[C/T]ATCTTTGGCG	S	C	T	C	C
G1095u5	WIAF-12539	HT2629	3048	KCNMA1, potassium large conductance calcium-activated channel, subfamily M, alpha member 1	CACTCATGAG[C/T]GCGACGTACT	S	C	T	S	S
G1095u6	WIAF-12544	HT2629	2352	KCNMA1, potassium large conductance calcium-activated channel, subfamily M, alpha member 1	GGATGTTTCA[C/T]TGGTGTGCAC	S	C	T	H	H
G1095u7	WIAF-12545	HT2629	2392	KCNMA1, potassium large conductance calcium-activated channel, subfamily M, alpha member 1	CATCCTGACT[C/T]GAAGTGAAGC	N	C	T	R	*

G1095u8	WIAF-12546	HT2629	2295	KCNMA1, potassium large conductance calcium-activated channel, subfamily M, alpha member 1	CTGGCAATGA[T/C]CAGATTGACA	S	T	C	D	D
G1095u9	WIAF-12548	HT2629	2949	KCNMA1, potassium large conductance calcium-activated channel, subfamily M, alpha member 1	AGTTTTTGGA[C/T]CAAGACGATG	S	C	T	D	D
G1095u10	WIAF-12549	HT2629	2865	KCNMA1, potassium large conductance calcium-activated channel, subfamily M, alpha member 1	TGCACGGGAT[G/A]TTACGTCAAC	M	G	A	M	I
G1096u1	WIAF-12547	L26318	930	PRKM8, protein kinase mitogen-activated 8 (MAP kinase)	TGCTGGTAAT[A/T]GATGCATCTA	S	A	T	I	I
G1098u1	WIAF-12515	L19711	2650	DAG1, dystroglycan 1 (dystrophin-associated glycoprotein 1)	TCTACCTGCA[C/T]ACAGTCATTC	S	C	T	H	H
G110u1	WIAF-10385	HT27392	230	meiosis-specific recA homolog, HsLim15	CAAAGGTATA[C/T]AGATGACAAC	N	C	T	Q	*
G110u2	WIAF-10397	HT27392	1050	meiosis-specific recA homolog, HsLim15	CCTGAAAATG[A/G]AGCCACCTTC	M	A	G	E	G
G110u3	WIAF-10399	HT27392	674	meiosis-specific recA homolog, HsLim15	TGAACATCAG[A/G]TGGAGCTACT	M	A	G	M	V
G1106u1	WIAF-12647	HT5073	5781	MAP1B, microtubule-associated protein 1B	ACTATGAGAA[G/A]ATAGAGAGAA	S	G	A	K	K
G1106u2	WIAF-12648	HT5073	5916	MAP1B, microtubule-associated protein 1B	CTGAAGAGGG[C/T]GGGTACTCAT	S	C	T	G	G
G1106u3	WIAF-12650	HT5073	1837	MAP1B, microtubule-associated protein 1B	AGACAAGCCA[G/A]TAAAAACAGA	M	G	A	V	I
G1106u4	WIAF-12653	HT5073	2476	MAP1B, microtubule-associated protein 1B	CACCACAGCA[G/A]CTGTCATGGC	M	G	A	A	T
G1106u5	WIAF-12656	HT5073	3913	MAP1B, microtubule-associated protein 1B	GCCCAATGAG[A/G]TTAAAGTCTC	M	A	G	I	V
G1106u6	WIAF-12667	HT5073	559	MAP1B, microtubule-associated protein 1B	GATTTTCACC[G/A]ATCAAGAGAT	M	G	A	D	N

G1106u7	WIAF-12668	HT5073	570	MAP1B, microtubule-associated protein 1B	ATCAAGAGAT[C/T]GGGGAGTTAC	S	C	T	I	I
G1106u8	WIAF-12669	HT5073	6175	MAP1B, microtubule-associated protein 1B	TACTTCCACA[T/C]ACTGTTACGA	M	T	C	Y	H
G1106u9	WIAF-12670	HT5073	1215	MAP1B, microtubule-associated protein 1B	TCACTCTCCA[G/C]TACCTAAACA	M	G	C	Q	H
G1106u10	WIAF-12672	HT5073	1821	MAP1B, microtubule-associated protein 1B	AGGTAATGGT[G/A]AAAAAAGACA	S	G	A	V	V
G1106u11	WIAF-12673	HT5073	2727	MAP1B, microtubule-associated protein 1B	GTCCTGCCGA[G/T]TCCCCTGATG	M	G	T	E	D
G1106u12	WIAF-12674	HT5073	2739	MAP1B, microtubule-associated protein 1B	CCCCTGATGA[G/A]GGAATCACTA	S	G	A	E	E
G1106u13	WIAF-12676	HT5073	3643	MAP1B, microtubule-associated protein 1B	AGATGCCACT[G/A]ATGGCAAGGA	M	G	A	D	N
G1106u14	WIAF-12677	HT5073	3609	MAP1B, microtubule-associated protein 1B	CACCGCTCAA[C/T]GGATTTTCTG	S	C	T	N	N
G1106u15	WIAF-12682	HT5073	4752	MAP1B, microtubule-associated protein 1B	TTCCAGAGCC[A/T]ACAACAGATG	S	A	T	P	P
G1110u1	WIAF-12517	HT1096	1527	myelin associated glycoprotein	GCGGCCTCGT[G/C]CTCACCAGCA	S	G	C	V	V
G1110u2	WIAF-12518	HT1096	1678	myelin associated glycoprotein	TGTGGGCGCC[G/T]TGGTCGCCTT	M	G	T	V	L
G1110u3	WIAF-12522	HT1096	1271	myelin associated glycoprotein	GCCGTGTCAC[C/T]CGAGGATGAT	M	C	T	P	L
G1113u1	WIAF-12523	HT2242	353	myelin transcription factor 1	AATTCCGATC[G/T]GATCCTCAGG	M	G	T	R	L
G1116a1	WIAF-13217	HT28451	417	myelin oligodendrocyte glycoprotein (MOG)	CAAGCTTATC[G/A]AGACCCTCTC	S	G	A	S	S
G1116a2	WIAF-13219	HT28451	913	myelin oligodendrocyte glycoprotein (MOG)	GCAGATCACT[C/G]TTGGCCTCGT	M	C	G	L	V
G1116a3	WIAF-13220	HT28451	922	myelin oligodendrocyte glycoprotein (MOG)	TCTTGGCCTC[G/A]TCTTCCTCTG	M	G	A	V	I
G1120u1	WIAF-12525	HT3695	1200	neurofilament, subunit H	TAGAGATAGC[T/C]GCTTACAGAA	S	T	C	A	A
G1123u1	WIAF-12542	HT2569	2269	OMG, oligodendrocyte myelin glycoprotein	CAGCTGCAAC[T/C]CTAACTATTC	S	T	C	T	T
G1126u1	WIAF-12526	HT28354	626	PSEN2, presenilin 2 (Alzheimer disease 4)	GAGCGAAGCA[T/C]GTGATCATGC	S	T	C	H	H
G1126u2	WIAF-12527	HT28354	494	PSEN2, presenilin 2 (Alzheimer disease 4)	ATGGAGAGAA[T/C]ACTGCCCAGT	S	T	C	N	N

G1126u3	WIAF-12528	HT28354	434	PSEN2, presenilin 2 (Alzheimer disease 4)	TAATGTCGGC[C/T]GAGAGCCCCA	S	C	T	A	A		
G1126u4	WIAF-12543	HT28354	550	PSEN2, presenilin 2 (Alzheimer disease 4)	GACCCTGACC[G/A]CTATGTCTGT	M	G	A	R	H		
G117u1	WIAF-10391	HT27765	156	GTBP, G/T mismatch-binding protein	ACTTCTCACC[A/G]GGAGATTTGG	S	A	G	P	P		
G117u2	WIAF-10392	HT27765	420	GTBP, G/T mismatch-binding protein	AACGTGCAGA[T/C]GAAGCCTTAA	S	T	C	D	D		
G117u3	WIAF-10407	HT27765	939	GTBP, G/T mismatch-binding protein	CCCACGTTAG[T/C]GGAGGTGGTG	S	T	C	S	S		
G117u4	WIAF-10411	HT27765	1622	GTBP, G/T mismatch-binding protein	CATTGTTCGA[G/A]ATTTAGGACT	M	G	A	R	K		
G117u5	WIAF-10412	HT27765	2405	GTBP, G/T mismatch-binding protein	GACAGCAGGG[C/T]TATAATGTAT	M	C	T	A	V		
G117u6	WIAF-10413	HT27765	2387	GTBP, G/T mismatch-binding protein	AAGAGTCAGA[A/T]CCACCCAGAC	M	A	T	N	I		
G125u1	WIAF-10371	HT28632	1999	ATM, ataxia telangiectasia mutated (includes complementation groups A, C and D)	CAGTAATTTT[C/T]CTCATCTTGT	M	C	T	P	S		
G125u2	WIAF-10372	HT28632	2631	ATM, ataxia telangiectasia mutated (includes complementation groups A, C and D)	TAATGAATGA[C/A]ATTGCAGATA	M	C	A	D	E		
G125u3	WIAF-10373	HT28632	3084	ATM, ataxia telangiectasia mutated (includes complementation groups A, C and D)	CAATGGAAGA[T/G]GTTCTTGAAC	M	T	G	D	E		
G125u5	WIAF-10375	HT28632	4767	ATM, ataxia telangiectasia mutated (includes complementation groups A, C and D)	CACTTATACC[C/T]CTTGTGTATG	S	C	T	P	P		
G125u6	WIAF-10383	HT28632	8713	ATM, ataxia telangiectasia mutated (includes complementation groups A, C and D)	ATTCTTGGAT[C/T]CAGCTATTTG	M	C	T	P	S		

G125u7	WIAF-10396	HT28632	1825	ATM, ataxia telangiectasia mutated (includes complementation groups A, C and D)	GACTTTGGCA[C/G]TGACCACCAG	M	C	G	L	V
G125u8	WIAF-10398	HT28632	2924	ATM, ataxia telangiectasia mutated (includes complementation groups A, C and D)	ACTACTGCTC[A/G]GACCAATACT	M	A	G	Q	R
G125u9	WIAF-10405	HT28632	8967	ATM, ataxia telangiectasia mutated (includes complementation groups A, C and D)	TTGAAGGTGT[C/T]TTCAGAAGAT	S	C	T	V	V
G125u10	WIAF-10408	HT28632	6954	ATM, ataxia telangiectasia mutated (includes complementation groups A, C and D)	CCAAACACCT[T/C]GTAGAACTCT	S	T	C	L	L
G125u11	WIAF-10409	HT28632	6855	ATM, ataxia telangiectasia mutated (includes complementation groups A, C and D)	TTCAGGAGCC[T/C]ATCATGGCTC	S	T	C	P	P
G125u12	WIAF-10410	HT28632	6801	ATM, ataxia telangiectasia mutated (includes complementation groups A, C and D)	TATATATTAA[G/T]TGGCAGAAAC	M	G	T	K	N
G125u13	WIAF-10421	HT28632	335	ATM, ataxia telangiectasia mutated (includes complementation groups A, C and D)	CATTCAGATT[C/G]CAAACAAGGA	M	C	G	S	C
G125u14	WIAF-11607	HT28632	3966	ATM, ataxia telangiectasia mutated (includes complementation groups A, C and D)	TTCCACATCT[G/A]GTGATTAGAA	S	G	A	L	L
G125a15	WIAF-13130	HT28632	8642	ATM, ataxia telangiectasia mutated (includes complementation groups A, C and D)	GAGAAATATG[A/C]AGTCTTCATG	M	A	C	E	A
G136u1	WIAF-10388	HT3337	535	MLH1, mutL (E. coli) homolog 1 (colon cancer, nonpolyposis type 2)	AGGAGAAAAG[C/T]TTTAAAAAAT	M	C	T	A	V

G136u2	WIAF-10389	HT3337	769	MLH1, mutL (E. coli) homolog 1 (colon cancer, nonpolyposis type 2)	TTCAAAATGA[A/G]TGGTTACATA	M	A	G	N	S
G144u1	WIAF-11638	HT3625	1129	FOS, v-fos FBJ murine osteosarcoma viral oncogene homolog	CCTGTGCACT[C/T]CGGTGGTCAC	M	C	T	P	S
G1461u1	WIAF-12562	HT0329	684	pRB-binding protein	TTGCCAAGAA[G/A]TCCAAGAACC	S	G	A	K	K
G1466u1	WIAF-12571	HT27849	2128	API2, apoptosis inhibitor 2	ATGATCCATG[G/C]GTAGAACATG	M	G	C	W	C
G1468u1	WIAF-12563	HT4986	1928	apoptosis inhibitor, neuronal	CCACCAGACC[A/T]GACGAGGGGC	S	A	T	P	P
G1468u2	WIAF-12564	HT4986	3057	apoptosis inhibitor, neuronal	TTTGCAATTC[C/G]TTCAAGGGAG	M	C	G	L	V
G1472u1	WIAF-12565	HT28478	242	BAK1, BCL2-antagonist/killer 1	GGCAGGAGTG[C/T]GGAGAGCCTG	S	C	T	C	C
G1472u2	WIAF-12572	HT28478	509	BAK1, BCL2-antagonist/killer 1	TGCAGCCCAC[G/A]GCAGAGAATG	S	G	A	T	T
G1473u1	WIAF-12568	HT28606	394	CASP6, caspase 6, apoptosis-related cysteine protease	GGTGTCAACT[G/C]TTAGCCACGC	M	G	C	V	L
G1473u2	WIAF-12576	HT28606	411	CASP6, caspase 6, apoptosis-related cysteine protease	ACGCAGATGC[C/T]GATTGCTTTG	S	C	T	A	A
G1479u1	WIAF-12550	Y09077	711	ATR, ataxia telangiectasia and Rad3 related	ACTTTATTAA[T/C]GGTTCTTACT	M	T	C	M	T
G1479u2	WIAF-12551	Y09077	4303	ATR, ataxia telangiectasia and Rad3 related	TTGCGTATGC[T/C]GATAATAGCC	S	T	C	A	A
G1479u3	WIAF-12552	Y09077	1894	ATR, ataxia telangiectasia and Rad3 related	ATTCTGATGA[T/C]GGCTGTTTAA	S	T	C	D	D
G1479u4	WIAF-12553	Y09077	1855	ATR, ataxia telangiectasia and Rad3 related	ATTTATGTGG[T/A]ATGCTCTCAC	S	T	A	G	G
G1479u5	WIAF-12558	Y09077	5287	ATR, ataxia telangiectasia and Rad3 related	TCATTCATTA[T/C]CATGGTGTAG	S	T	C	Y	Y

G1479u6	WIAF-12559	Y09077	5539	ATR, ataxia telangiectasia and Rad3 related	CAGCTTTTTA[T/C]GACTCACTGA	S	T	C	Y	Y
G1479u7	WIAF-12569	Y09077	1540	ATR, ataxia telangiectasia and Rad3 related	ATCCTGTTAT[T/C]GAGATGTTAG	S	T	C	I	I
G1479u8	WIAF-12570	Y09077	2521	ATR, ataxia telangiectasia and Rad3 related	ATTTAATGGA[A/G]GATCCAGACA	S	A	G	E	E
G1482u1	WIAF-12560	HT27870	3176	BLM, Bloom syndrome	AAAATATAAC[G/A]GAATGCAGGA	S	G	A	T	T
G1482u2	WIAF-12561	HT27870	3605	BLM, Bloom syndrome	GAAATAAAGC[C/A]CAAACTGTAC	S	C	A	A	A
G1482u3	WIAF-12573	HT27870	2677	BLM, Bloom syndrome	TATGTATTAC[C/T]GAAAAAGCCT	M	C	T	P	L
G1483u1	WIAF-12597	HT1470	1910	MYBL2, v-myb avian myeloblastosis viral oncogene homolog-like 2	GGATGAGGAT[G/A]TGAAGCTGAT	M	G	A	V	M
G1483u2	WIAF-12610	HT1470	244	MYBL2, v-myb avian myeloblastosis viral oncogene homolog-like 2	ATGAGGAGGA[C/T]GAGCAGCTGA	S	C	T	D	D
G1483u3	WIAF-12611	HT1470	1406	MYBL2, v-myb avian myeloblastosis viral oncogene homolog-like 2	CACTGAGAAT[A/G]GCACCAGTCT	M	A	G	S	G
G1485u1	WIAF-12581	HT1432	1941	BCR, breakpoint cluster region	TGGAGATGAG[A/G]AAATGGGTCC	S	A	G	R	R
G1485u2	WIAF-12582	HT1432	3144	BCR, breakpoint cluster region	TGACCATCAA[T/C]AAGGAAGATG	S	T	C	N	N
G1485u3	WIAF-12583	HT1432	3777	BCR, breakpoint cluster region	ATAACAAGGA[T/C]GTGTCGGTGA	S	T	C	D	D
G1485u4	WIAF-12603	HT1432	2831	BCR, breakpoint cluster region	CAGATCAAGA[G/A]TGACATCCAG	M	G	A	S	N
G1485u5	WIAF-12608	HT1432	4217	BCR, breakpoint cluster region	ATCCCTGCCC[C/T]GGACAGCAAG	M	C	T	P	L
G1486u1	WIAF-12578	HT33770	1909	BRCA2, breast cancer 2, early onset	ATTGATAATG[G/A]AAGCTGGCCA	M	G	A	G	E
G1486u2	WIAF-12579	HT33770	3623	BRCA2, breast cancer 2, early onset	AGTTTAGAAA[A/G]CCAAGCTACA	S	A	G	K	K
G1486u3	WIAF-12586	HT33770	1341	BRCA2, breast cancer 2, early onset	AAATGTAGCA[A/C]ATCAGAAGCC	M	A	C	N	H
G1486u4	WIAF-12594	HT33770	446	BRCA2, breast cancer 2, early onset	CTTATAATCA[G/A]CTGGCTTCAA	S	G	A	Q	Q

G1486u5	WIAF-12598	HT33770	3013	BRCA2, breast cancer 2, early onset	ACCATGGTTT [T/C] ATATGGAGAC	M	T	C	L	S	
G1486u6	WIAF-12599	HT33770	3187	BRCA2, breast cancer 2, early onset	GAAAAAAATA [A/T] TGATTACATG	M	A	T	N	I	
G1486u7	WIAF-12604	HT33770	4971	BRCA2, breast cancer 2, early onset	AGCATGTGAG [A/C] CCATTGAGAT	M	A	C	T	P	
G1486u8	WIAF-12607	HT33770	4034	BRCA2, breast cancer 2, early onset	ATGATTCTGT [C/T] GTTTCAATGT	S	C	T	V	V	
G1487u1	WIAF-12584	HT27632	2536	BRCA1, breast cancer 1, early onset	AGTCAGTGTG [C/G] AGCATTTGAA	M	C	G	A	G	
G1487u2	WIAF-12587	HT27632	4697	BRCA1, breast cancer 1, early onset	CATCTCAAGA [G/C] GAGCTCATTA	M	G	C	E	D	
G1487u3	WIAF-12595	HT27632	469	BRCA1, breast cancer 1, early onset	TCTCCTGAAC [A/G] TCTAAAAGAT	M	A	G	H	R	
G1487u4	WIAF-12600	HT27632	3667	BRCA1, breast cancer 1, early onset	AGCGTCCAGA [A/G] AGGAGAGCTT	M	A	G	K	R	
G1487u5	WIAF-12601	HT27632	3537	BRCA1, breast cancer 1, early onset	TATGGGAAGT [A/G] GTCATGCATC	M	A	G	S	G	
G1487u6	WIAF-12602	HT27632	4956	BRCA1, breast cancer 1, early onset	ATCTGCCCAG [A/G] GTCCAGCTGC	M	A	G	S	G	
G1487u7	WIAF-12605	HT27632	2090	BRCA1, breast cancer 1, early onset	AGTACAACCA [A/G] ATGCCAGTCA	S	A	G	Q	Q	
G1487u8	WIAF-12614	HT27632	233	BRCA1, breast cancer 1, early onset	TCTCCACAAA [G/A] TGTGACCACA	S	G	A	K	K	
G1492u1	WIAF-12585	HT3506	3912	cell death-associated kinase	TCCAGGTCCG [T/C] GGCCTGGAGA	S	T	C	R	R	
G1492u2	WIAF-12593	HT3506	4352	cell death-associated kinase	TACAACACCA [A/G] TAACGGGGCT	M	A	G	N	S	
G1492u3	WIAF-12606	HT3506	2127	cell death-associated kinase	GCAATTTGGA [C/T] ATCTCCAACA	S	C	T	D	D	
G1492u4	WIAF-12612	HT3506	1605	cell death-associated kinase	TGAAATTTCT [C/T] AGTGAGAACA	S	C	T	L	L	
G1494u1	WIAF-12589	HT28507	366	cell death-inducing protein Bik	TTCACCACAC [T/C] TAAGGAGAAC	M	T	C	L	P	
G1495u1	WIAF-12580	HT27803	759	CSE1L, chromosome segregation 1 (yeast homolog)-like	TTTCTTCCCT [G/C] ATCCTGATCT	S	G	C	L	L	
G1501u1	WIAF-13502	HT1949	1181	MCC, mutated in colorectal cancers	CAGCAATGAC [A/C] TTCCCATCGC	M	A	C	I	L	

G1501u2	WIAF-13503	HT1949	1753	MCC, mutated in colorectal cancers	CAGCTGAGAA[C/T]GCTGCCAAGG	S	C	T	N	N
G1501u3	WIAF-13504	HT1949	2344	MCC, mutated in colorectal cancers	TGTCCCTAGC[T/C]GAACTCAGGA	S	T	C	A	A
G1501u4	WIAF-13521	HT1949	445	MCC, mutated in colorectal cancers	AGCGAACGAC[G/A]CTTCGCTATG	S	G	A	T	T
G1501u5	WIAF-13522	HT1949	1504	MCC, mutated in colorectal cancers	AAAGCAATGC[T/C]GAGAGGATGA	S	T	C	A	A
G1501u6	WIAF-13527	HT1949	2511	MCC, mutated in colorectal cancers	TTCGTGAATG[A/G]TCTAAAGCGG	M	A	G	D	G
G1502u1	WIAF-12633	HT1547	870	CCND1, cyclin D1 (PRAD1: parathyroid adenomatosis 1)	AGTGTGACCC[A/G]GACTGCCTCC	S	A	G	P	P
G1503u1	WIAF-13741	U37022	1151	CDK4, cyclin-dependent kinase 4	CATGCCAATT[G/A]CATCGTTCAC	M	G	A	C	Y
G1503u2	WIAF-13742	U37022	1410	CDK4, cyclin-dependent kinase 4	CTGAAGCCGA[C/T]CAGTTGGGCA	S	C	T	D	D
G1503u3	WIAF-13743	U37022	1328	CDK4, cyclin-dependent kinase 4	TATGCAACAC[C/T]TGTGGACATG	M	C	T	P	L
G1503u4	WIAF-13780	U37022	1194	CDK4, cyclin-dependent kinase 4	TTCTGGTGAC[A/G]AGTGGTGGAA	S	A	G	T	T
G1503u5	WIAF-13781	U37022	1443	CDK4, cyclin-dependent kinase 4	TGATTGGGCT[G/A]CCTCCAGAGG	S	G	A	L	L
G1503u6	WIAF-13787	U37022	1633	CDK4, cyclin-dependent kinase 4	CTCTTATCTA[C/T]ATAAGGATGA	M	C	T	H	Y
G1517u1	WIAF-12618	HT1132	3894	ERBB3, v-erb-b2 avian erythroblastic leukemia viral oncogene homolog 3	CAGACCTCAG[T/C]GCCTCTCTGG	S	T	C	S	S
G152u1	WIAF-11608	HT3854	1673	HSPA1L, heat shock 70kD protein-like 1	GTGAGTGATG[A/C]AGGTTTGAAG	M	A	C	E	A
G152u2	WIAF-11629	HT3854	1683	HSPA1L, heat shock 70kD protein-like 1	AAGGTTTGAA[G/A]GGCAAGATTA	S	G	A	K	K
G152u3	WIAF-11609	HT3854	1478	HSPA1L, heat shock 70kD protein-like 1	GTCACAGCCA[C/T]GGACAAGAGC	M	C	T	T	M
G152u4	WIAF-11610	HT3854	1443	HSPA1L, heat shock 70kD protein-like 1	TGACGTTTGA[C/T]ATTGATGCCA	S	C	T	D	D
G1520u1	WIAF-12162	HT1175	2211	DNA excision repair protein ERCC2, 5' end	TGACCGTGGA[C/T]GAGGGTGTCC	S	C	T	D	D

G1520u2	WIAF-12166	HT1175	546	DNA excision repair protein ERCC2, 5' end	CCCACTGCCG[A/C]TTCTATGAGG	S	A	C	R	R
G1527u1	WIAF-12168	HT0086	577	GSTM2, glutathione S-transferase M2 (muscle)	TCATCTCCCG[A/C]TTTGAGGGCT	S	A	C	R	R
G1527u2	WIAF-12169	HT0086	644	GSTM2, glutathione S-transferase M2 (muscle)	ACCTGTGTTC[A/T]CAAAGATGGC	M	A	T	T	S
G1527u3	WIAF-12171	HT0086	100	GSTM2, glutathione S-transferase M2 (muscle)	ACTCAAGCTA[C/T]GAGGAAAAGA	S	C	T	Y	Y
G1527u4	WIAF-12172	HT0086	41	GSTM2, glutathione S-transferase M2 (muscle)	GGGGTACTGG[A/G]ACATCCGCGG	M	A	G	N	D
G1527u5	WIAF-12173	HT0086	215	GSTM2, glutathione S-transferase M2 (muscle)	GATTGATGGG[A/G]CTCACAAGAT	M	A	G	T	A
G1527u6	WIAF-12194	HT0086	238	GSTM2, glutathione S-transferase M2 (muscle)	CCCAGAGCAA[T/C]GCCATCCTGC	S	T	C	N	N
G1528u1	WIAF-11950	HT1811	529	GSTM3, glutathione S-transferase M3 (brain)	GTATATTTGA[C/G]CCCAAGTGCC	M	C	G	D	E
G1528u2	WIAF-11951	HT1811	674	GSTM3, glutathione S-transferase M3 (brain)	CAACAAGCCT[G/A]TATGCTGAGC	M	G	A	V	I
G1528u3	WIAF-11989	HT1811	572	GSTM3, glutathione S-transferase M3 (brain)	GGCTTTCATG[T/G]GCCGTTTTGA	M	T	G	C	G
G1528u4	WIAF-13470	HT1811	240	GSTM3, glutathione S-transferase M3 (brain)	CAGAGCAATG[C/A]CATCTTGCGC	M	C	A	A	D
G1529u1	WIAF-14146	HT2006	797	GSTM4, glutathione S-transferase M4	TGGACGCCTT[C/T]CCAAATCTGA	S	C	T	F	F
G153u1	WIAF-12163	HT3856	1212	HSPA1B, heat shock 70kD protein 1	TGGGGCTGGA[G/A]ACGGCCGGAG	S	G	A	E	E
G153u2	WIAF-12182	HT3856	676	HSPA1B, heat shock 70kD protein 1	GGCCGGGGAC[A/G]CCCACCTGGG	M	A	G	T	A
G153u3	WIAF-12183	HT3856	1695	HSPA1B, heat shock 70kD protein 1	TCAGCGAGGC[C/G]GACAAGAAGA	S	C	G	A	A
G153u4	WIAF-12189	HT3856	330	HSPA1B, heat shock 70kD protein 1	ACAAGGGGGA[G/C]ACCAAGGCAT	M	G	C	E	D
G153u5	WIAF-12190	HT3856	1053	HSPA1B, heat shock 70kD protein 1	AGCTGCTGCA[A/G]GACTTCTTCA	S	A	G	Q	Q
G1530u1	WIAF-11964	HT3010	673	GSTM5, glutathione S-transferase M5	ATTCCTCCGA[G/A]GTCTTTTGTT	M	G	A	G	S
G1530u2	WIAF-11995	HT3010	593	GSTM5, glutathione S-transferase M5	GACGCCTTCC[T/C]AAACTTGAAG	M	T	C	L	P

G1530u3	WIAF-13473	HT3010	693	GSTM5, glutathione S-transferase M5	TTGGAAAGTC[A/G]GCTACATGGA	S	A	G	S	S
G1533u1	WIAF-13458	HT27460	543	GSTT2, glutathione S-transferase theta 2	CTCTCGGCTA[C/T]GAACTGTTTG	S	C	T	Y	Y
G1533u2	WIAF-13460	HT27460	417	GSTT2, glutathione S-transferase theta 2	GGACTGCCAT[G/A]GACCAGGCCC	M	G	A	M	I
G1533u3	WIAF-13461	HT27460	359	GSTT2, glutathione S-transferase theta 2	CAGGTGTTGG[G/A]GCCACTCATT	M	G	A	G	E
G1533u4	WIAF-13462	HT27460	363	GSTT2, glutathione S-transferase theta 2	TGTTGGGGCC[A/C]CTCATTGGGG	S	A	C	P	P
G1533u5	WIAF-13463	HT27460	385	GSTT2, glutathione S-transferase theta 2	CCAGGTGCCC[G/A]AGGAGAAGGT	M	G	A	E	K
G1535u1	WIAF-11952	HT0436	517	HCK, hemopoietic cell kinase	CCGCGTTGAC[T/C]CTCTGGAGAC	M	T	C	S	P
G1535u2	WIAF-12013	HT0436	783	HCK, hemopoietic cell kinase	TGGACCACTA[C/T]AAGAAGGGGA	S	C	T	Y	Y
G1535u3	WIAF-13464	HT0436	357	HCK, hemopoietic cell kinase	TCATCGTGGT[T/C]GCCCTGTATG	S	T	C	V	V
G1535u4	WIAF-13465	HT0436	387	HCK, hemopoietic cell kinase	CCATTCACCA[C/T]GAAGACCTCA	S	C	T	H	H
G1535u5	WIAF-13466	HT0436	471	HCK, hemopoietic cell kinase	CCCTGGCCAC[C/G]CGGAAGGAGG	S	C	G	T	T
G1535u6	WIAF-13467	HT0436	240	HCK, hemopoietic cell kinase	CCAGCGCCAG[C/T]CCACACTGTC	S	C	T	S	S
G1535u7	WIAF-13468	HT0436	394	HCK, hemopoietic cell kinase	CCACGAAGAC[C/T]TCAGCTTCCA	M	C	T	L	F
G1537u1	WIAF-12020	U04045	1514	MSH2, mutS (E. coli) homolog 2 (colon cancer, nonpolyposis type 1)	GTGAATTAAG[A/G]GAAATAATGA	S	A	G	R	R
G1537u2	WIAF-12044	U04045	599	MSH2, mutS (E. coli) homolog 2 (colon cancer, nonpolyposis type 1)	GACTGTGTGA[A/T]TTCCCTGATA	M	A	T	E	D
G1537u3	WIAF-12045	U04045	1452	MSH2, mutS (E. coli) homolog 2 (colon cancer, nonpolyposis type 1)	AGATATGGAT[C/T]AGGTGGAAAA	N	C	T	Q	*
G1537u4	WIAF-12076	U04045	938	MSH2, mutS (E. coli) homolog 2 (colon cancer, nonpolyposis type 1)	GACAGTTTGA[A/T]CTGACTACTT	M	A	T	E	D

G1537u5	WIAF-12077	U04045	1878	MSH2, mutS (E. coli) homolog 2 (colon cancer, nonpolyposis type 1)	TCAGCTAGAT[G/A]CTGTTGTCAG	M	G	A	A	T
G1543u1	WIAF-13856	J00119	553	MOS, v-mos Moloney murine sarcoma viral oncogene homolog	GAGTTTCTGG[G/T]CTGAGCTCAA	M	G	T	A	S
G1543u2	WIAF-13857	J00119	621	MOS, v-mos Moloney murine sarcoma viral oncogene homolog	GCACGCGCAC[G/A]CCCGCAGGGT	S	G	A	T	T
G1544u1	WIAF-12018	U59464	3821	PTCH, patched (Drosophila) homolog	CATCCCGAAT[C/T]CAGGCATCAC	M	C	T	S	F
G1544u2	WIAF-12019	U59464	3618	PTCH, patched (Drosophila) homolog	GCGTGGTCCG[C/T]TTCGCCATGC	S	C	T	R	R
G1544u3	WIAF-12027	U59464	1761	PTCH, patched (Drosophila) homolog	ATTTTGCCAT[G/T]GTTCTGCTCA	M	G	T	M	I
G1544u4	WIAF-12029	U59464	4074	PTCH, patched (Drosophila) homolog	CTGCCATGGG[C/T]AGCTCCGTGC	S	C	T	G	G
G1544u5	WIAF-12043	U59464	3845	PTCH, patched (Drosophila) homolog	CCCTCGAACC[C/T]GAGACAGCAG	M	C	T	P	L
G1544u6	WIAF-12056	U59464	1433	PTCH, patched (Drosophila) homolog	CTGCTGGTTG[C/T]ACTGTCAGTG	M	C	T	A	V
G1544u7	WIAF-12058	U59464	3298	PTCH, patched (Drosophila) homolog	CACCGTTCAC[G/C]TTGCTTTGGC	M	G	C	V	L
G1544u8	WIAF-12062	U59464	3986	PTCH, patched (Drosophila) homolog	TCTACTGAAG[G/A]GCATTCTGGC	M	G	A	G	E
G1544u9	WIAF-13489	U59464	1665	PTCH, patched (Drosophila) homolog	CCATCAGCAA[T/C]GTCACAGCCT	S	T	C	N	N
G1544u10	WIAF-13490	U59464	2396	PTCH, patched (Drosophila) homolog	AAATACTTTT[C/T]TTTCTACAAC	M	C	T	S	F
G1544u11	WIAF-13491	U59464	2199	PTCH, patched (Drosophila) homolog	GGACACTCTC[A/G]TCTTTTGCTG	S	A	G	S	S
G1544u12	WIAF-13492	U59464	2222	PTCH, patched (Drosophila) homolog	AAGCACTATG[C/T]TCCTTTCCTC	M	C	T	A	V
G1544u13	WIAF-13500	U59464	1686	PTCH, patched (Drosophila) homolog	TCTTCATGGC[C/T]GCGTTAATCC	S	C	T	A	A
G1545u1	WIAF-12032	HT0473	1835	RAG1, recombination activating gene 1	GGACATGGAA[G/A]AAGACATCTT	M	G	A	E	K
G1545u2	WIAF-12035	HT0473	2519	RAG1, recombination activating gene 1	TGACATTGGC[A/G]ATGCAGCTGA	M	A	G	N	D

G1545u3	WIAF-12046	HT0473	3045	RAG1, recombination activating gene 1	CGGAAAATGA[A/G]TGCCAGGCAG	M	A	G	N	S
G1545u4	WIAF-12047	HT0473	3146	RAG1, recombination activating gene 1	TCATAATGCA[T/C]TAAAAACCTC	S	T	C	L	L
G1545u5	WIAF-12075	HT0473	2513	RAG1, recombination activating gene 1	CCACTGTGAC[A/T]TTGGCAATGC	M	A	T	I	F
G1545u6	WIAF-13484	HT0473	1322	RAG1, recombination activating gene 1	GTCGCTGACT[C/T]GGAGAGCTCA	M	C	T	R	W
G1545u7	WIAF-13494	HT0473	2571	RAG1, recombination activating gene 1	GAAGTGTATA[A/G]GAATCCCAAT	M	A	G	K	R
G1545u8	WIAF-13498	HT0473	1018	RAG1, recombination activating gene 1	TTCTGGCTGA[C/A]CCTGTGGAGA	M	C	A	D	E
G1545u9	WIAF-13499	HT0473	2782	RAG1, recombination activating gene 1	ATCTTTACCT[G/C]AAGATGAAAC	S	G	C	L	L
G1548u1	WIAF-12015	HT4999	133	IFI27, interferon, alpha-inducible protein 27	CTCTGCCGTA[G/A]TTTTGCCCCT	M	G	A	V	I
G1548u2	WIAF-13482	HT4999	380	IFI27, interferon, alpha-inducible protein 27	ATCCTGGGCT[C/T]CATTGGGTCT	M	C	T	S	F
G1548u3	WIAF-13483	HT4999	135	IFI27, interferon, alpha-inducible protein 27	CTGCCGTAGT[T/C]TTGCCCCTGG	S	T	C	V	V
G155u1	WIAF-11634	HT3962	991	CHC1, chromosome condensation 1	AGCTGGATGT[G/A]CCTGTGGTAA	S	G	A	V	V
G155u2	WIAF-11635	HT3962	1271	CHC1, chromosome condensation 1	CGGCTTCGGC[C/T]TCTCCAACTA	M	C	T	L	F
G155u3	WIAF-11636	HT3962	1192	CHC1, chromosome condensation 1	GCCGGGGCCA[C/T]GTGAGATTCC	S	C	T	H	H.
G155u4	WIAF-11637	HT3962	1267	CHC1, chromosome condensation 1	TGTACGGCTT[C/T]GGCCTCTCCA	S	C	T	F	F
G155u5	WIAF-11649	HT3962	1657	CHC1, chromosome condensation 1	TGATGGGCAA[A/G]CAGCTGGAGA	S	A	G	K	K
G1550u1	WIAF-12057	M16038	611	LYN, v-yes-1 Yamaguchi sarcoma viral related oncogene homolog	GCAAAGTCCC[T/G]TTTAACAAAA	M	T	G	L	R
G1550u2	WIAF-12061	M16038	1371	LYN, v-yes-1 Yamaguchi sarcoma viral related oncogene homolog	TGGCATACAT[C/T]GAGCGGAAGA	S	C	T	I	I
G1550u3	WIAF-12080	M16038	1059	LYN, v-yes-1 Yamaguchi sarcoma viral related oncogene homolog	AAAGGCTTGG[C/T]GCTGGGCAGT	S	C	T	G	G

G1550u4	WIAF-12081	M16038	996	LYN, v-yes-1 Yamaguchi sarcoma viral related oncogene homolog	AGCCACAGAA[G/A]CCATGGGATA	S	G	A	K	K
G1552u1	WIAF-12030	HT4578	2355	PMS1, postmeiotic segregation increased (S. cerevisiae) 1	CCTGCTATTT[A/T]AAAGACTTCT	N	A	T	K	*
G1552u2	WIAF-12031	HT4578	2231	PMS1, postmeiotic segregation increased (S. cerevisiae) 1	ACAAAGTTGA[C/T]TTAGAAGAGA	S	C	T	D	D
G1552u3	WIAF-12040	HT4578	617	PMS1, postmeiotic segregation increased (S. cerevisiae) 1	TCATGAGCTT[T/C]GGTATCCTTA	S	T	C	F	F
G1552u4	WIAF-12063	HT4578	1723	PMS1, postmeiotic segregation increased (S. cerevisiae) 1	TCATGTAACA[A/G]AAAATCAAAT	M	A	G	K	R
G1552u5	WIAF-12064	HT4578	1732	PMS1, postmeiotic segregation increased (S. cerevisiae) 1	AAAAAATCAA[A/G]TGTAATAGAT	M	A	G	N	S
G1552u6	WIAF-12065	HT4578	1660	PMS1, postmeiotic segregation increased (S. cerevisiae) 1	TTACCATGTA[A/G]AGTAAGTAAT	M	A	G	K	R
G1552u7	WIAF-12066	HT4578	1975	PMS1, postmeiotic segregation increased (S. cerevisiae) 1	GAACGATACA[A/G]TAGTCAAATG	M	A	G	N	S
G1552u8	WIAF-12067	HT4578	1881	PMS1, postmeiotic segregation increased (S. cerevisiae) 1	TTTAGAGGAT[G/T]CAACACTACA	M	G	T	A	S
G1552u9	WIAF-12068	HT4578	2454	PMS1, postmeiotic segregation increased (S. cerevisiae) 1	TTTAGACGTT[T/A]TATATAAAAT	M	T	A	L	I
G1552u10	WIAF-12069	HT4578	2457	PMS1, postmeiotic segregation increased (S. cerevisiae) 1	AGACGTTTTA[T/C]ATAAAATGAC	M	T	C	Y	H
G1552u11	WIAF-12082	HT4578	2557	PMS1, postmeiotic segregation increased (S. cerevisiae) 1	ATACCAGGAG[T/C]TTCAATTACT	M	T	C	V	A
G1552u12	WIAF-12083	HT4578	971	PMS1, postmeiotic segregation increased (S. cerevisiae) 1	TTTTCTTTCT[G/T]AAAATCGATG	S	G	T	L	L

G1554u1	WIAF-12028	HT4161	1500	ELK3, ELK3, ETS-domain protein (SRF accessory protein 2) NOTE: Symbol and name provisional.	CTCAGAAATC[C/T]TGATGACGTC	S	C	T	S	S
G1554u2	WIAF-12059	HT4161	1380	ELK3, ELK3, ETS-domain protein (SRF accessory protein 2) NOTE: Symbol and name provisional.	CTGCCAGGCT[G/A]CAAGGGCCAA	S	G	A	L	L
G1554u3	WIAF-12060	HT4161	1436	ELK3, ELK3, ETS-domain protein (SRF accessory protein 2) NOTE: Symbol and name provisional.	CACATGCCAG[T/C]GCCAATCCCC	M	T	C	V	A
G1562u1	WIAF-12024	HT28220	804	PDCD1, programmed cell death 1	GGGGCTCAGC[T/C]GACGGCCCTC	S	T	C	A	A
G1562u2	WIAF-13488	HT28220	644	PDCD1, programmed cell death 1	GACCCCTCAG[C/T]CGTGCCTGTG	M	C	T	A	V
G1563u1	WIAF-13493	HT1187	1748	EGFR, epidermal growth factor receptor (avian erythroblastic leukemia viral (v-erb-b) oncogene homolog)	CCGGAGCCCA[G/A]GGACTGCGTC	M	G	A	R	K
G1563u2	WIAF-13497	HT1187	2073	EGFR, epidermal growth factor receptor (avian erythroblastic leukemia viral (v-erb-b) oncogene homolog)	ACGGATGCAC[T/A]GGGCCAGGTC	S	T	A	T	T
G1566u1	WIAF-12016	HT27594	235	PDCD2, programmed cell death 2	GCGCCGCTGC[C/G]TGGCCGCCCG	M	C	G	P	R
G1566u2	WIAF-12033	HT27594	904	PDCD2, programmed cell death 2	TTGGAATTCC[A/G]GGTCATGCCT	M	A	G	Q	R
G1566u3	WIAF-12041	HT27594	331	PDCD2, programmed cell death 2	AATCAACTAC[C/T]CAGGAAAAAC	M	C	T	P	L
G1566u4	WIAF-12071	HT27594	649	PDCD2, programmed cell death 2	CCTGAGGTTG[T/C]GGAAAAGGAA	M	T	C	V	A
G1566u5	WIAF-12072	HT27594	633	PDCD2, programmed cell death 2	AGAAGATGAG[A/T]TTATGCCTGA	M	A	T	I	F
G1567u1	WIAF-12042	M95936	293	AKT2, v-akt murine thymoma viral oncogene homolog 2	GAGAGGCCGC[G/A]ACCCAACACC	M	G	A	R	Q

G1572u1	WIAF-12212	HT3998	1894	proto-oncogene c-abl, tyrosine protein kinase, alt. transcript 2	TGTTCCAGGA[A/G]TCCAGTATCT	S	A	G	E				E
G1572u2	WIAF-12233	HT3998	3694	proto-oncogene c-abl, tyrosine protein kinase, alt. transcript 2	AGCTTCAGAT[C/T]TGCCCGGCGA	S	C	T	I				I
G1572u3	WIAF-12234	HT3998	3721	proto-oncogene c-abl, tyrosine protein kinase, alt. transcript 2	GCAGTGGTCC[G/A]GCGGCCACTC	S	G	A	P				P
G1573u1	WIAF-12021	HT0642	343	CBL, Cas-Br-M (murine) ecotropic retroviral transforming sequence	TCATGGACAA[G/C]GTGGTGCGGT	M	G	C	K				N
G1573u2	WIAF-12022	HT0642	363	CBL, Cas-Br-M (murine) ecotropic retroviral transforming sequence	TTGTGTCAGA[A/T]CCCAAAGCTG	M	A	T	N				I
G1573u3	WIAF-12034	HT0642	2364	CBL, Cas-Br-M (murine) ecotropic retroviral transforming sequence	AATATTCAGT[C/T]CCAGGCGCCA	M	C	T	S				F
G1573u4	WIAF-12049	HT0642	387	CBL, Cas-Br-M (murine) ecotropic retroviral transforming sequence	CTAAAGAATA[G/A]CCCACCTTAT	M	G	A	S				N
G1573u5	WIAF-12050	HT0642	947	CBL, Cas-Br-M (murine) ecotropic retroviral transforming sequence	AACTCATCCT[G/A]GCTACATGGC	M	G	A	G				S
G1573u6	WIAF-12070	HT0642	2740	CBL, Cas-Br-M (murine) ecotropic retroviral transforming sequence	TCGAGAACCT[C/T]ATGAGTCAGG	S	C	T	L				L
G1573u7	WIAF-12073	HT0642	661	CBL, Cas-Br-M (murine) ecotropic retroviral transforming sequence	TCTTTCCAAG[T/C]GGACTCTTTC	S	T	C	S				S
G1573u8	WIAF-12074	HT0642	2569	CBL, Cas-Br-M (murine) ecotropic retroviral transforming sequence	CTCTGGATGG[T/C]GATCCTACAA	S	T	C	G				G
G1573u9	WIAF-13486	HT0642	2006	CBL, Cas-Br-M (murine) ecotropic retroviral transforming sequence	CCGGCACTCA[C/T]TTCCATTTTC	M	C	T	L				F

G1574u1	WIAF-12037	HT1508	2493	FES, feline sarcoma (Snyder-Theilen) viral (v-fes)/Fujinami avian sarcoma (PRCII) viral (v-fps) oncogene homolog	AGCGGCCCAG[C/T]TTCAGCACCA	S	C	T	S	S
G1574u2	WIAF-12051	HT1508	189	FES, feline sarcoma (Snyder-Theilen) viral (v-fes)/Fujinami avian sarcoma (PRCII) viral (v-fps) oncogene homolog	CCCAGCGGGT[C/T]AAGAGTGACA	S	C	T	V	V
G1574u3	WIAF-12052	HT1508	1441	FES, feline sarcoma (Snyder-Theilen) viral (v-fes)/Fujinami avian sarcoma (PRCII) viral (v-fps) oncogene homolog	GAAGCCCCTG[C/T]ATGAGCAGCT	M	C	T	H	Y
G1574u4	WIAF-12053	HT1508	2202	FES, feline sarcoma (Snyder-Theilen) viral (v-fes)/Fujinami avian sarcoma (PRCII) viral (v-fps) oncogene homolog	GAGAGGAAGC[C/T]GATGGGGTCT	S	C	T	A	A
G1574u5	WIAF-12054	HT1508	2088	FES, feline sarcoma (Snyder-Theilen) viral (v-fes)/Fujinami avian sarcoma (PRCII) viral (v-fps) oncogene homolog	CTGCTGGCAT[G/T]GAGTACCTGG	M	G	T	M	I
G1574u6	WIAF-12078	HT1508	1577	FES, feline sarcoma (Snyder-Theilen) viral (v-fes)/Fujinami avian sarcoma (PRCII) viral (v-fps) oncogene homolog	GATGGTCTGC[C/T]CCGGCACTTC	M	C	T	P	L
G1574u7	WIAF-13495	HT1508	579	FES, feline sarcoma (Snyder-Theilen) viral (v-fes)/Fujinami avian sarcoma (PRCII) viral (v-fps) oncogene homolog	GTGACAAGGC[T/C]AAGGACAAGT	S	T	C	A	A
G1575u1	WIAF-12079	HT1052	963	FGR, Gardner-Rasheed feline sarcoma viral (v-fgr) oncogene homolog	TGGGCACCGG[C/T]TGCTTCGGGG	S	C	T	G	G

G1575u2	WIAF-13487	HT1052	232	FGR, Gardner-Rasheed feline sarcoma viral (v-fgr) oncogene homolog	CAGAAGCTAC[G/A]GGGCAGCAGA	M	G	A	G	R
G1585u1	WIAF-12017	HT1675	996	CRK, v-crk avian sarcoma virus CT10 oncogene homolog	TGGATCAACA[G/A]AATCCCGATG	S	G	A	Q	Q
G1585u2	WIAF-12036	HT1675	446	CRK, v-crk avian sarcoma virus CT10 oncogene homolog	ACTACAACGT[T/C]GATAGAACCA	M	T	C	L	S
G1587u1	WIAF-12023	HT0590	1473	proto-oncogene dbl	GGCCAATCCA[A/G]TTTGTGGTAC	S	A	G	Q	Q
G1587u2	WIAF-12025	HT0590	2549	proto-oncogene dbl	GTCCAGGCTT[C/T]TAATGTAGAT	M	C	T	S	F
G1587u3	WIAF-12026	HT0590	2828	proto-oncogene dbl	GCATCACAAT[C/T]TGCAGAAATC	M	C	T	S	F
G1587u4	WIAF-12038	HT0590	982	proto-oncogene dbl	AAATTCTCAG[G/C]AGCTATTATC	M	G	C	E	Q
G1587u5	WIAF-12039	HT0590	2343	proto-oncogene dbl	AACCAATGCA[G/T]CGACACCTTT	M	G	T	Q	H
G1587u6	WIAF-12048	HT0590	683	proto-oncogene dbl	GACACTGAAG[G/A]AGCTGTCAGT	M	G	A	G	E
G1587u7	WIAF-12055	HT0590	2686	proto-oncogene dbl	TTCTCTTCAG[C/T]AGAATGATGA	N	C	T	Q	*
G1587u8	WIAF-13485	HT0590	2136	proto-oncogene dbl	ACTGTGAAGG[T/A]TCTGCTCTGT	S	T	A	G	G
G1587u9	WIAF-13496	HT0590	1566	proto-oncogene dbl	AAAATCAGAG[C/T]AACTTAAAAA	S	C	T	S	S
G159u1	WIAF-11616	HT4209	1059	RAD23B, RAD23 (S. cerevisiae) homolog B	AGTACTGGGG[C/T]TCCTCAGTCT	M	C	T	A	V
G1590u1	WIAF-13897	HT2455	1257	ETS2, v-ets avian erythroblastosis virus E26 oncogene homolog 2	GCCAGTCTCT[C/G]TGCCTCAATA	S	C	G	L	L
G1590u2	WIAF-13913	HT2455	1107	ETS2, v-ets avian erythroblastosis virus E26 oncogene homolog 2	ATTCTGGGAC[T/G]CCCAAAGACC	S	T	G	T	T
G1590u3	WIAF-13914	HT2455	1314	ETS2, v-ets avian erythroblastosis virus E26 oncogene homolog 2	GGAGTGACCC[A/G]GTGGAGCAAG	S	A	G	P	P
G1591u1	WIAF-13924	HT2333	417	HRAS, v-Ha-ras Harvey rat sarcoma viral oncogene homolog	TCCAGAACCA[T/C]TTTGTGGACG	S	T	C	H	H
G1595u1	WIAF-12262	HT33778	1302	proto-oncogene 1-myc, alt. transcript 1	GCATACCTCA[G/C]TGGCTACTAA	M	G	C	S	T
G1597u1	WIAF-12243	HT0410	900	MAS1, MAS1 oncogene	CCATCTTGGT[C/T]GTGAAGATCC	S	C	T	V	V
G160u1	WIAF-11630	HT4247	690	RAD23A, RAD23 (S. cerevisiae) homolog A	AGAGCCAGGT[A/G]TCGGAGCAGC	S	A	G	V	V
G1602u1	WIAF-14180	HT1903	1321	proto-oncogene pim-1	GTCGCCGGGG[C/A]CCAGCAAATA	M	C	A	P	T

G1604u1	WIAF-12319	HT2788	1182	REL, v-rel avian reticuloendotheliosis viral oncogene homolog	CCTCCCAAAG[T/C]GCTGGGATTA	S	T	C	S	S
G1609u1	WIAF-12358	HT33646	348	RIPK1, receptor (TNFRSF)-interacting serine-threonine kinase 1	GACGCAGGGT[C/T]TCCCATGACC	S	C	T	V	V
G161u1	WIAF-11654	HT4251	1522	DNA repair and recombination homolog RAD52	TATGATCCAT[C/T]TTAACTGAGG	M	C	T	S	F
G1610a1	WIAF-12101	HT27727	501	replication protein Rpa4, 30 kDa	TGCAACTCCT[G/A]CTATTAAGAC	M	G	A	A	T
G1610a2	WIAF-12102	HT27727	554	replication protein Rpa4, 30 kDa	TACCGTGTAA[C/T]GTGAACCAGC	S	C	T	N	N
G1610u3	WIAF-12307	HT27727	450	replication protein Rpa4, 30 kDa	TTCTGCTGCT[G/A]ATGGAGCGAG	M	G	A	D	N
G1610u4	WIAF-12320	HT27727	1037	replication protein Rpa4, 30 kDa	TGATTCATGA[G/C]TGTCCTCATC	M	G	C	E	D
G1610u5	WIAF-12321	HT27727	857	replication protein Rpa4, 30 kDa	TAGAGGACAT[G/A]AACGAGTTCA	M	G	A	M	I
G1610u6	WIAF-12343	HT27727	539	replication protein Rpa4, 30 kDa	GAATTCAGGA[C/T]GTTGTACCGT	S	C	T	D	D
G1630u1	WIAF-12302	HT3563	4312	DCC, deleted in colorectal carcinoma	ACTCATGAAG[C/T]AGCTTAATGC	N	C	T	Q	*
G1632u1	WIAF-13572	HT27355	742	tumor suppressor, PDGF receptor beta-like	TTTATGACAT[G/C]AAGCGGGGCT	M	G	C	M	I
G1632u2	WIAF-13584	HT27355	1102	tumor suppressor, PDGF receptor beta-like	TGGAAGACTT[C/T]GAGACGATTG	S	C	T	F	F
G1632u3	WIAF-13601	HT27355	258	tumor suppressor, PDGF receptor beta-like	AAGACGCAGT[C/T]TATCATGATG	M	C	T	S	F
G1633u1	WIAF-13957	HT1778	1263	FER, fer (fps/fes related) tyrosine kinase (phosphoprotein NCP94)	TTCAGGCAAA[T/C]GAGATCATGT	S	T	C	N	N
G1633u2	WIAF-13958	HT1778	2407	FER, fer (fps/fes related) tyrosine kinase (phosphoprotein NCP94)	TATGTTGTAT[C/T]TCGAGAGTAA	M	C	T	L	F
G1634u1	WIAF-13505	HT3216	1569	ELK1, ELK1, member of ETS oncogene family	TCTCGACCCC[C/T]GTGGTGCTCT	S	C	T	P	P
G1634u2	WIAF-13858	HT3216	456	ELK1, ELK1, member of ETS oncogene family	GGCTGTGGGG[A/G]CTACGCAAGA	S	A	G	G	G

G1634u3	WIAF-13859	HT3216	745	ELK1, ELK1, member of ETS oncogene family	AGGCCCAGGC[G/A]GTTTGGCACG	M	G	A	G	S
G1638u1	WIAF-14172	HT1224	98	uracil-DNA glycosylase	GCTGGGACCT[G/C]TTCCACAAAT	-	G	C	-	-
G1643u1	WIAF-13517	HT3751	629	DXS648E, DNA segment on chromosome X (unique) 648 expressed sequence	TACATCCCCA[G/A]TCGTGGCCCT	M	G	A	S	N
G1645u1	WIAF-14087	D21089	363	XPC, xeroderma pigmentosum, complementation group C	AAAACCTCAA[G/A]GTTATAAAGG	S	G	A	K	K
G1645u2	WIAF-14088	D21089	2166	XPC, xeroderma pigmentosum, complementation group C	TGCATTCCAG[G/A]GACACGTGGC	S	G	A	R	R
G1645u3	WIAF-14089	D21089	1580	XPC, xeroderma pigmentosum, complementation group C	GGGAGCCATC[G/A]TAAGGACCCA	M	G	A	R	H
G1645u4	WIAF-14090	D21089	1601	XPC, xeroderma pigmentosum, complementation group C	AGCTTGCCAG[T/C]GGCATCCTCA	M	T	C	V	A
G1645u5	WIAF-14091	D21089	2920	XPC, xeroderma pigmentosum, complementation group C	CCCATTTGAG[A/C]AGCTGTGAGC	M	A	C	K	Q
G1645u6	WIAF-14103	D21089	405	XPC, xeroderma pigmentosum, complementation group C	ATGACCTCAG[G/A]GACTTTCCAA	S	G	A	R	R
G1645u7	WIAF-14104	D21089	151	XPC, xeroderma pigmentosum, complementation group C	GGGACGCGAA[C/G]TGCGCAGCCA	M	C	G	L	V
G1645u8	WIAF-14105	D21089	2133	XPC, xeroderma pigmentosum, complementation group C	AAGCGGTCTA[C/T]TCCAGGGATT	S	C	T	Y	Y
G167u1	WIAF-11632	HT4579	83	PMS2L8, postmeiotic segregation increased 2-like 8	CCTATTGATC[G/A]GAAGTCAGTC	M	G	A	R	Q
G167u2	WIAF-11633	HT4579	219	PMS2L8, postmeiotic segregation increased 2-like 8	GAGTGGATCT[T/C]ATTGAAGTTT	S	T	C	L	L
G167u3	WIAF-11644	HT4579	768	PMS2L8, postmeiotic segregation increased 2-like 8	TGCCCCCTAG[T/C]GACTCCGTGT	S	T	C	S	S

G167u4	WIAF-11622	HT4579	1645	PMS2L8, postmeiotic segregation increased 2-like 8	GAAAGCGCCT[G/A]AAACTGACGA	M	G	A	E	K
G167u5	WIAF-11645	HT4579	1512	PMS2L8, postmeiotic segregation increased 2-like 8	ACTCGGGGCA[C/T]GGCAGCACTT	S	C	T	H	H
G167u6	WIAF-11646	HT4579	1619	PMS2L8, postmeiotic segregation increased 2-like 8	TCGCAGGAAC[A/G]TGTGGACTCT	M	A	G	H	R
G167u7	WIAF-11647	HT4579	1432	PMS2L8, postmeiotic segregation increased 2-like 8	CGTCCTGAGA[C/T]CTCAGAAAGA	M	C	T	P	S
G167u8	WIAF-11625	HT4579	2490	PMS2L8, postmeiotic segregation increased 2-like 8	GGACTGCTCT[T/C]AACACAAGCG	S	T	C	L	L
G167u9	WIAF-11619	HT4579	804	PMS2L8, postmeiotic segregation increased 2-like 8	TGAGCTGTTC[G/C]GATGCTCTGC	S	G	C	S	S
G167u10	WIAF-11623	HT4579	1555	PMS2L8, postmeiotic segregation increased 2-like 8	CATCCCAGAC[A/G]CGGGCAGTCA	M	A	G	T	A
G167u11	WIAF-11624	HT4579	2364	PMS2L8, postmeiotic segregation increased 2-like 8	CCTTCGGACC[C/T]CAGGACGTCG	S	C	T	P	P
G167u12	WIAF-11626	HT4579	2348	PMS2L8, postmeiotic segregation increased 2-like 8	ACTAGTAAAA[A/G]CTGGACCTTC	M	A	G	N	S
G181u1	WIAF-11697	HT48793	311	ERCC4, excision repair cross-complementing rodent repair deficiency, complementation group 4	ATATTTGCGA[C/T]AAGTAGGATA	M	C	T	T	I
G181u2	WIAF-11698	HT48793	295	ERCC4, excision repair cross-complementing rodent repair deficiency, complementation group 4	CACACAAGGT[G/C]GTGTTATATT	M	G	C	G	R
G181u3	WIAF-11699	HT48793	234	ERCC4, excision repair cross-complementing rodent repair deficiency, complementation group 4	TTGAACACCT[C/T]CCTCGCCGTG	S	C	T	L	L

G181u4	WIAF-11704	HT48793	808	ERCC4, excision repair cross-complementing rodent repair deficiency, complementation group 4	TTTGTGGCAC[C/T]AGCTTGGAGC	N	C	T	Q	*
G181u5	WIAF-11705	HT48793	640	ERCC4, excision repair cross-complementing rodent repair deficiency, complementation group 4	TTCTATGACA[C/T]CTACCATGCT	M	C	T	P	S
G181u6	WIAF-11670	HT48793	1117	ERCC4, excision repair cross-complementing rodent repair deficiency, complementation group 4	AGAAAGCAAC[C/T]CAAAGTGGGA	M	C	T	P	S
G185u1	WIAF-11668	HT5122	319	ACVR2B, activin A receptor, type IIB	TCTGCAACGA[G/A]CGCTTCACTC	S	G	A	E	E
G185u2	WIAF-11707	HT5122	70	ACVR2B, activin A receptor, type IIB	AGACACGGGA[G/C]TGCATCTACT	M	G	C	E	D
G185u3	WIAF-11672	HT5122	812	ACVR2B, activin A receptor, type IIB	CCTCACGGAT[T/C]ACCTCAAGGG	M	T	C	Y	H
G185u4	WIAF-13542	X77533	1109	ACVR2B, activin A receptor, type IIB	GGCTCCTGAG[G/A]TGCTCGAGGG	M	G	A	V	M
G185u5	WIAF-13558	X77533	997	ACVR2B, activin A receptor, type IIB	TGCTGAAGAG[C/T]GACCTCACAG	S	C	T	S	S
G187u1	WIAF-11669	HT97400	183	androgen	CCAGAGACAG[C/T]GCGACCCGGA	M	C	T	R	C
G191u1	WIAF-10176	AF025375	414	CXCR4, chemokine (C-X-C motif), receptor 4 (fusin)	ACCTGGCCAT[C/T]GTCCACGCCA	S	C	T	I	I
G193u1	WIAF-10178	D29984	231	CCR2, chemokine (C-C motif) receptor 2	AGTGCTTGAC[T/A]GACATTTACC	S	T	A	T	T
G193u2	WIAF-10179	D29984	190	CCR2, chemokine (C-C motif) receptor 2	CATGCTGGTC[G/A]TCCTCATCTT	M	G	A	V	I
G194u1	WIAF-10211	D43767	121	SCYA17, small inducible cytokine subfamily A (Cys-Cys), member 17	ACATCCACGC[A/C]GCTCGAGGGA	S	A	C	A	A
G197u1	WIAF-10167	D50403	1515	NRAMP1, natural resistance-associated macrophage protein 1 (might include Leishmaniasis)	GGTGCTAGTC[T/C]GCGCCATCAA	M	T	C	C	R

G197u2	WIAF-10173	D50403	1629	NRAMP1, natural resistance-associated macrophage protein 1 (might include Leishmaniasis)	CACCTACCTG[G/C]TCTGGACCTG	M	G	C	V	L
G20u1	WIAF-10249	U14722	896	ACVR1B, activin A receptor, type IB	CGGTACACAG[T/C]GACAATTGAG	M	T	C	V	A
G20u2	WIAF-10250	U14722	866	ACVR1B, activin A receptor, type IB	GAGCACGGGT[C/T]CCTGTTTGAT	M	C	T	S	F
G20u3	WIAF-10251	U14722	1391	ACVR1B, activin A receptor, type IB	CAGAGTTATG[A/T]GGCACTGCGG	M	A	T	E	V
G20u4	WIAF-10252	U14722	1236	ACVR1B, activin A receptor, type IB	TATATTGGGA[G/C]ATTGCTCGAA	M	G	C	E	D
G20u5	WIAF-10261	U14722	518	ACVR1B, activin A receptor, type IB	GAGATGTGTC[T/C]CTCCAAAGAC	M	T	C	L	P
G207a1	WIAF-10516	L25259	866	Human CTLA4 counter-receptor (B7-2) mRNA, complete cds.	AGCTGTACTT[C/T]CAACAGTTAT	M	C	T	P	S
G208u1	WIAF-10204	L31581	85	CCR7, chemokine (C-C motif) receptor 7	GGGGAAACCA[A/G]TGAAAAGCGT	M	A	G	M	V
G211u1	WIAF-10213	M24545	174	SCYA2, small inducible cytokine A2 (monocyte chemotactic protein 1, homologous to mouse Sig-je)	TCACCTGCTG[T/C]TATAACTTCA	S	T	C	C	C
G214u1	WIAF-10191	M27533	452	CD80, CD80 antigen (CD28 antigen ligand 1, B7-1 antigen)	TGAAAGAAGT[G/A]GCAACGCTGT	S	G	A	V	V
G215u1	WIAF-11659	M28393	822	PRF1, perforin 1 (preforming protein)	GCATCTCTGC[C/T]GAAGCCAAGG	S	C	T	A	A
G215u2	WIAF-11723	M28393	159	PRF1, perforin 1 (preforming protein)	TGACCAGCCT[C/T]CGCCGCTCGG	S	C	T	L	L
G215u3	WIAF-11724	M28393	96	PRF1, perforin 1 (preforming protein)	CAGAGTGCAA[G/A]CGCAGCCACA	S	G	A	K	K
G215u4	WIAF-11725	M28393	1377	PRF1, perforin 1 (preforming protein)	ATAACAACCC[C/T]ATCTGGTCAG	S	C	T	P	P
G215u5	WIAF-11726	M28393	1326	PRF1, perforin 1 (preforming protein)	TGAAGCTCTT[C/T]TTTGGTGGCC	S	C	T	F	F

G215u6	WIAF-11727	M28393	1076	PRF1, perforin 1 (preforming protein)	CGGCGGGAGG[C/T]ACTGAGGAGG	M	C	T	A	V
G217u1	WIAF-11691	M31932	649	FCGR2B, Fc fragment of IgG, low affinity IIb, receptor for (CD32)	GCAGCTCTTC[A/G]CCAATGGGGA	S	A	G	S	S
G217u2	WIAF-11692	M31932	625	FCGR2B, Fc fragment of IgG, low affinity IIb, receptor for (CD32)	TCACTGTCCA[A/G]GTGCCCAGCA	S	A	G	Q	Q
G217u3	WIAF-11712	M31932	332	FCGR2B, Fc fragment of IgG, low affinity IIb, receptor for (CD32)	GACTGGCCAG[A/C]CCAGCCTCAG	M	A	C	T	P
G217u4	WIAF-11713	M31932	101	FCGR2B, Fc fragment of IgG, low affinity IIb, receptor for (CD32)	GGCTTCTGCA[G/T]ACAGTCAAGC	M	G	T	D	Y
G218u1	WIAF-10184	M36712	677	CD8B1, CD8 antigen, beta polypeptide 1 (p37)	TTTTACAAAT[A/G]AGCAGAGAAT	N	A	G	*	*
G218u2	WIAF-10188	M36712	326	CD8B1, CD8 antigen, beta polypeptide 1 (p37)	GCTGTGTTTC[G/C]GGATGCAAGC	M	G	C	R	P
G218u3	WIAF-10189	M36712	196	CD8B1, CD8 antigen, beta polypeptide 1 (p37)	CAGTAACATG[C/T]GCATCTACTG	M	C	T	R	C
G218u4	WIAF-10190	M36712	225	CD8B1, CD8 antigen, beta polypeptide 1 (p37)	AGCGCCAGGC[A/C]CCGAGCAGTG	S	A	C	A	A
G218u5	WIAF-10194	M36712	583	CD8B1, CD8 antigen, beta polypeptide 1 (p37)	GGTGGCTGGC[G/A]TCCTGGTTCT	M	G	A	V	I
G218u6	WIAF-10208	M36712	372	CD8B1, CD8 antigen, beta polypeptide 1 (p37)	TGAAGCCGGA[A/G]GACAGTGGCA	S	A	G	E	E
G218u7	WIAF-10209	M36712	400	CD8B1, CD8 antigen, beta polypeptide 1 (p37)	CTGCATGATC[G/T]TCGGGAGCCC	M	G	T	V	F
G218u8	WIAF-10210	M36712	270	CD8B1, CD8 antigen, beta polypeptide 1 (p37)	TCTGGGATTC[C/T]GCAAAAGGGA	S	C	T	S	S
G218a9	WIAF-10518	M36712	618	CD8B1, CD8 antigen, beta polypeptide 1 (p37)	GAGTGGCCAT[C/G]CACCTGTGCT	M	C	G	I	M
G218a10	WIAF-13223	M36712	556	CD8B1, CD8 antigen, beta polypeptide 1 (p37)	TTGTAGCCCC[A/G]TCACCCTTGG	M	A	G	I	V
G218a11	WIAF-13224	M36712	836	CD8B1, CD8 antigen, beta polypeptide 1 (p37)	CTGTGTGTGA[T/C]GTGCATGGGA	-	T	C	-	-
G22u1	WIAF-10301	U86136	6719	Human telomerase-associated protein TP-1 mRNA, complete cds.	GGTGGTAACC[G/A]TCGGGCTAGA	M	G	A	V	I

G22u2	WIAF-10302	U86136	7537	Human telomerase-associated protein TP-1 mRNA, complete cds.	CTGATGGGAT[C/G]CTATGGAACC	M	C	G	I	M
G22u3	WIAF-10311	U86136	1798	Human telomerase-associated protein TP-1 mRNA, complete cds.	ATGATGCCAT[T/C]GATGCCCTCG	S	T	C	I	I
G22u4	WIAF-10312	U86136	2397	Human telomerase-associated protein TP-1 mRNA, complete cds.	CTGTCTCTGG[C/T]TGGCCAAAGG	M	C	T	A	V
G22u5	WIAF-10313	U86136	3289	Human telomerase-associated protein TP-1 mRNA, complete cds.	AGAAAGGGAT[A/C]ACCTGCCGCA	S	A	C	I	I
G22u6	WIAF-10314	U86136	3242	Human telomerase-associated protein TP-1 mRNA, complete cds.	AGAGGCCGCA[T/C]GTCGGATCTC	M	T	C	C	R
G22u7	WIAF-10315	U86136	4482	Human telomerase-associated protein TP-1 mRNA, complete cds.	CCGTTTGCCT[G/A]CCTCGTCCAG	M	G	A	C	Y
G22u8	WIAF-10316	U86136	4363	Human telomerase-associated protein TP-1 mRNA, complete cds.	GTTTGACTGT[G/A]GACCAGCTGC	S	G	A	V	V
G22u9	WIAF-10317	U86136	4230	Human telomerase-associated protein TP-1 mRNA, complete cds.	GTGTCTGAGA[G/A]ACTCCGGACC	M	G	A	R	K
G22u10	WIAF-10318	U86136	4419	Human telomerase-associated protein TP-1 mRNA, complete cds.	GGGACTAAGA[G/C]CTGGGAAGAA	M	G	C	S	T
G22u11	WIAF-10319	U86136	5269	Human telomerase-associated protein TP-1 mRNA, complete cds.	TCTCCGATGA[T/C]ACACTCTTTC	S	T	C	D	D
G22u12	WIAF-10320	U86136	5015	Human telomerase-associated protein TP-1 mRNA, complete cds.	GCTGCTCTCC[C/T]GGAGATGGCA	M	C	T	R	W
G22u13	WIAF-10321	U86136	5133	Human telomerase-associated protein TP-1 mRNA, complete cds.	GTGGCCTTCT[C/T]CACCAATGGG	M	C	T	S	F
G22u14	WIAF-10322	U86136	7764	Human telomerase-associated protein TP-1 mRNA, complete cds.	ACAGCCCTCC[A/G]TGTGCTACCT	M	A	G	H	R

G650u1	WIAF-12323	X52773	862	RXRA, retinoid X receptor, alpha	CTCGCCGAAC[G/A]ACCCTGTCAC	M	G	A	D	N
G650u2	WIAF-12341	X52773	102	RXRA, retinoid X receptor, alpha	TCCTGCCGCT[C/T]GATTTCTCCA	S	C	T	L	L
G650u3	WIAF-12348	X52773	673	RXRA, retinoid X receptor, alpha	GGCCATGGGC[A/G]TGAAGCGGGA	M	A	G	M	V
G650u4	WIAF-12349	X52773	902	RXRA, retinoid X receptor, alpha	GACAAACAGC[T/C]TTTCACCCTG	M	T	C	L	P
G653a1	WIAF-13326	HT1458	439	RARB, retinoic acid receptor, beta	AGGAGAAAGC[T/C]CTCAAAGCAT	S	T	C	A	A
G655a1	WIAF-13327	J05252	1158	PCSK2, proprotein convertase subtilisin/kexin type 2	CCTTCAGCAA[C/T]GGGAGGAAAA	S	C	T	N	N
G655a2	WIAF-13334	J05252	678	PCSK2, proprotein convertase subtilisin/kexin type 2	CCTATCCTTA[C/A]CCTCGGTACA	N	C	A	Y	*
G655a3	WIAF-13335	J05252	744	PCSK2, proprotein convertase subtilisin/kexin type 2	TTTCTGCTGC[C/T]GCCAACAACA	S	C	T	A	A
G658u1	WIAF-11856	J02943	971	CBG, corticosteroid binding globulin	TCTATGACCT[T/C]GGAGATGTGC	S	T	C	L	L
G658u2	WIAF-13407	J02943	771	CBG, corticosteroid binding globulin	CCTTCATGAC[T/G]CAGAGCTCCC	M	T	G	S	A
G658u3	WIAF-13408	J02943	773	CBG, corticosteroid binding globulin	TTCATGACTC[A/G]GAGCTCCCCT	S	A	G	S	S
G658u4	WIAF-13409	J02943	1046	CBG, corticosteroid binding globulin	TCACCCAGGA[C/T]GCCCAGCTGA	S	C	T	D	D
G663u1	WIAF-13400	HT3157	1202	TPO, thyroid peroxidase	CGCCACGCGC[G/A]CCTGCGGCCT	S	G	A	A	A
G663u2	WIAF-13401	HT3157	1282	TPO, thyroid peroxidase	GGCCGCCGCCCA[G/C]CGAGGTCCCC	M	G	C	S	T
G668a1	WIAF-13350	U53506	350	DIO2, deiodinase, iodothyronine, type II	TCGATGCCTA[C/A]AACAGGTGA	N	C	A	Y	*
G668a2	WIAF-13351	U53506	354	DIO2, deiodinase, iodothyronine, type II	TGCCTACAAA[C/A]AGGTGAAATT	M	C	A	Q	K
G668a3	WIAF-13352	U53506	408	DIO2, deiodinase, iodothyronine, type II	TGTCTCCAGT[A/G]CAGAAGGAGG	M	A	G	T	A
G673a1	WIAF-13328	M57464	1723	Human ret proto-oncogene mRNA for tyrosine kinase.	CGAGCCTGGG[G/A]AGCCCCGGGG	M	G	A	E	K
G673a2	WIAF-13336	M57464	1186	Human ret proto-oncogene mRNA for tyrosine kinase.	GGCTCGCCGA[T/A]TTGCCCAGAT	M	T	A	F	I

G673a3	WIAF-13337	M57464	1227	Human ret proto-oncogene mRNA for tyrosine kinase.	ACTGCCAGGC[G/A]TTCAGTGGCA	S	G	A	A	A
G673a4	WIAF-13338	M57464	2118	Human ret proto-oncogene mRNA for tyrosine kinase.	TTGGAAAAAC[T/A]CTAGGAGAAG	S	T	A	T	T
G673a5	WIAF-13339	M57464	2238	Human ret proto-oncogene mRNA for tyrosine kinase.	CGAGTGAGCT[T/G]CGAGACCTGC	S	T	G	L	L
G678a1	WIAF-13353	D49492	1439	GDF10, growth differentiation factor 10	TCGGCTGGAA[T/A]GAATGGATAA	M	T	A	N	K
G68u1	WIAF-10434	HT1115	1214	ERCC3, excision repair cross-complementing rodent repair deficiency, complementation group 3 (xeroderma pigmentosum group B complementing)	CTGTGGAGCA[G/A]TGGAAAGCCC	S	G	A	Q	Q
G68u2	WIAF-10435	HT1115	1155	ERCC3, excision repair cross-complementing rodent repair deficiency, complementation group 3 (xeroderma pigmentosum group B complementing)	TGTGACTGCT[G/C]CATGCACTGT	M	G	C	A	P
G68u3	WIAF-10436	HT1115	1327	ERCC3, excision repair cross-complementing rodent repair deficiency, complementation group 3 (xeroderma pigmentosum group B complementing)	AGCACCTACT[C/T]CATGCTGGGC	M	C	T	S	F
G68u4	WIAF-10461	HT1115	926	ERCC3, excision repair cross-complementing rodent repair deficiency, complementation group 3 (xeroderma pigmentosum group B complementing)	AGGAAATGAT[T/C]GAGGAACTCC	S	T	C	I	I
G68u5	WIAF-10464	HT1115	1430	ERCC3, excision repair cross-complementing rodent repair deficiency, complementation group 3 (xeroderma pigmentosum group B complementing)	AAGTGCACAC[C/T]ATACCAGCCA	S	C	T	T	T

G684a1	WIAF-13359	X51801	712	BMP7, bone morphogenetic protein 7 (osteogenic protein 1)	GTTTATCAGG[T/G]GCTCCAGGAG	M	T	G	V	G
G684a2	WIAF-13360	X51801	719	BMP7, bone morphogenetic protein 7 (osteogenic protein 1)	AGGTGCTCCA[G/A]GAGCACTTGG	S	G	A	Q	Q
G684a3	WIAF-13361	X51801	796	BMP7, bone morphogenetic protein 7 (osteogenic protein 1)	GGCTGGCTGG[T/G]GTTTGACATC	M	T	G	V	G
G684a4	WIAF-13362	X51801	862	BMP7, bone morphogenetic protein 7 (osteogenic protein 1)	GGCCTGCAGC[T/G]CTCGGTGGAG	M	T	G	L	R
G684a5	WIAF-13363	X51801	658	BMP7, bone morphogenetic protein 7 (osteogenic protein 1)	ATCTACAAGG[A/G]CTACATCCGG	M	A	G	D	G
G684u6	WIAF-13834	X51801	1421	BMP7, bone morphogenetic protein 7 (osteogenic protein 1)	GCCACTAGCT[C/T]CTCCGAGAAT	-	C	T	-	-
G685a1	WIAF-13329	D89675	882	BMPR1B, bone morphogenetic protein receptor, type IB	GTTCCCTTTA[T/G]GATTATCTGA	N	T	G	Y	*
G685a2	WIAF-13330	D89675	920	BMPR1B, bone morphogenetic protein receptor, type IB	GCTAAATCAA[T/C]GCTGAAGTTA	M	T	C	M	T
G685a3	WIAF-13331	D89675	770	BMPR1B, bone morphogenetic protein receptor, type IB	TATCAGACAG[T/G]GTTGATGAGG	M	T	G	V	G
G685a4	WIAF-13340	D89675	1303	BMPR1B, bone morphogenetic protein receptor, type IB	TCCTTATCAT[G/A]ACCTAGTGCC	M	G	A	D	N
G685a5	WIAF-13341	D89675	1372	BMPR1B, bone morphogenetic protein receptor, type IB	GTTACGCCCC[T/G]CATTCCCAAA	M	T	G	S	A
G685a6	WIAF-13342	D89675	1173	BMPR1B, bone morphogenetic protein receptor, type IB	TGTTGGACGA[G/A]AGCTTGAACA	S	G	A	E	E
G686u1	WIAF-13816	Z48923	2705	BMPR2, bone morphogenetic protein receptor, type II (serine/threonine kinase)	AAATTTGGCA[G/A]CAAGCACAAA	M	G	A	S	N

G686u2	WIAF-13817	Z48923	2749	BMPR2, bone morphogenetic protein receptor, type II (serine/threonine kinase)	TGGAGTTGCC[A/T]AGATGAATAC	N	A	T	K	*
G687a1	WIAF-13343	HT1455	626	CALB1, calbindin 1, (28kD)	ATGATCAGGA[C/T]GGCAATGGAT	S	C	T	D	D
G696u1	WIAF-11839	HT27700	1075	calcium-sensing receptor	GGGCACAATT[G/C]CAGCTGATGA	M	G	C	A	P
G696u2	WIAF-11840	HT27700	1551	calcium-sensing receptor	TACCTGTGGA[C/T]ACCTTTCTGA	S	C	T	D	D
G696u3	WIAF-11841	HT27700	1688	calcium-sensing receptor	TTACGGATAT[C/T]CTACAATGTG	M	C	T	S	F
G696u4	WIAF-11842	HT27700	1698	calcium-sensing receptor	CCTACAATGT[G/T]TACTTAGCAG	S	G	T	V	V
G696u5	WIAF-11858	HT27700	1767	calcium-sensing receptor	GGAGAGGGCT[C/T]TTCACCAATG	S	C	T	L	L
G696u6	WIAF-11859	HT27700	1689	calcium-sensing receptor	TACGGATATC[C/T]TACAATGTGT	S	C	T	S	S
G696u7	WIAF-11860	HT27700	2541	calcium-sensing receptor	TCGTGCTCTG[C/T]ATCTCATGCA	S	C	T	C	C
G696u8	WIAF-11861	HT27700	2581	calcium-sensing receptor	TGTCCTCCTG[G/A]TGTTTGAGGC	M	G	A	V	M
G696u9	WIAF-11863	HT27700	3159	calcium-sensing receptor	TCTCCCGCAA[G/C]CGGTCCAGCA	M	G	C	K	N
G696u10	WIAF-11872	HT27700	562	calcium-sensing receptor	TCCTATTCAT[T/A]TTGGAGTAGC	M	T	A	F	I
G696u11	WIAF-11878	HT27700	2941	calcium-sensing receptor	CATTCCAGCC[T/G]ATGCCAGCAC	M	T	G	Y	D
G696u12	WIAF-13386	HT27700	1145	calcium-sensing receptor	AGGGATATCT[G/A]CATCGACTTC	M	G	A	C	Y
G696u13	WIAF-13395	HT27700	670	calcium-sensing receptor	GATATTTGCC[A/G]TAGAGGAGAT	M	A	G	I	V
G696u14	WIAF-13396	HT27700	2243	calcium-sensing receptor	TTCTGGTCCA[A/G]TGAGAACCAC	M	A	G	N	S
G696u15	WIAF-13397	HT27700	2742	calcium-sensing receptor	AGCTGGAGGA[T/C]GAGATCATCT	S	T	C	D	D
G698u1	WIAF-13547	X61598	393	CBP1, collagen-binding protein 1	TCAGCAACTC[G/C]ACGGCGCGCA	S	G	C	S	S
G698u2	WIAF-13549	X61598	628	CBP1, collagen-binding protein 1	CGGCGCCCTG[C/T]TAGTCAACGC	S	C	T	L	L
G698u3	WIAF-13550	X61598	1230	CBP1, collagen-binding protein 1	GCGGCTCCCT[G/A]CTATTCATTG	S	G	A	L	L
G701u1	WIAF-12382	HT27657	706	CGRP type I receptor	AACGATGTTG[C/A]AGCAGGAACT	M	C	A	A	E
G701u2	WIAF-12391	HT27657	841	CGRP type I receptor	TGGACAAATT[A/T]TACCCAGTGT	M	A	T	Y	F
G704u1	WIAF-14046	X60382	1396	COL10A1, collagen, type X, alpha 1 (Schmid metaphyseal chondrodysplasia)	AGGCATTCCA[G/A]GATTCCCTGG	M	G	A	G	R
G704u2	WIAF-14070	X60382	1648	COL10A1, collagen, type X, alpha 1 (Schmid metaphyseal chondrodysplasia)	TGCCAACCAG[G/C]GGGTAACAGG	M	G	C	G	R

G704u3	WIAF-14071	X60382	1824	COL10A1, collagen, type X, alpha 1 (Schmid metaphyseal chondrodysplasia)	CATACCACGT [G/C] CATGTGAAAG	S	G	C	V	V
G704u4	WIAF-14072	X60382	1582	COL10A1, collagen, type X, alpha 1 (Schmid metaphyseal chondrodysplasia)	AGTCATGCCT [G/C] AGGGTTTTAT	M	G	C	E	Q
G705a1	WIAF-13228	J04177	686	COL11A1, collagen, type XI, alpha 1	AGAAGAAAAC [T/A] GTGACAATGA	S	T	A	T	T
G705a2	WIAF-13229	J04177	698	COL11A1, collagen, type XI, alpha 1	TGACAATGAT [T/A] GTTGATTGTA	S	T	A	I	I
G705a3	WIAF-13230	J04177	888	COL11A1, collagen, type XI, alpha 1	TAGTCCAGAC [T/A] GTGACTCTTC	M	T	A	C	S
G705a4	WIAF-13231	J04177	894	COL11A1, collagen, type XI, alpha 1	AGACTGTGAC [T/A] CTTCAGCACC	M	T	A	S	T
G705a5	WIAF-13232	J04177	651	COL11A1, collagen, type XI, alpha 1	TGACGGGAAG [T/A] GGCATCGGGT	M	T	A	W	R
G705a6	WIAF-13233	J04177	661	COL11A1, collagen, type XI, alpha 1	TGGCATCGGG [T/A] AGCAATCAGC	M	T	A	V	E
G705a7	WIAF-13234	J04177	1597	COL11A1, collagen, type XI, alpha 1	CGTCCTGGCT [T/C] ACCAGGGGCT	M	T	C	L	S
G705a8	WIAF-13235	J04177	2745	COL11A1, collagen, type XI, alpha 1	TGGGTTTCCA [G/A] GTGCCAATGG	M	G	A	G	S
G705a9	WIAF-13236	J04177	4385	COL11A1, collagen, type XI, alpha 1	GTCCAGAAGG [T/A] CTTCGGGGCA	S	T	A	G	G
G705a10	WIAF-13237	J04177	4576	COL11A1, collagen, type XI, alpha 1	GAAAAAGGTG [A/T] CCGAGGGCTC	M	A	T	D	V
G705a11	WIAF-13238	J04177	4306	COL11A1, collagen, type XI, alpha 1	GCTAAGGGGG [A/C] AGCAGGTGCA	M	A	C	E	A
G705a12	WIAF-13239	J04177	4837	COL11A1, collagen, type XI, alpha 1	AGACATACTG [A/G] AGGCATGCAA	M	A	G	E	G
G705a13	WIAF-13240	J04177	4931	COL11A1, collagen, type XI, alpha 1	AACAAGACAT [C/T] GAGCATATGA	S	C	T	I	I
G705a14	WIAF-13346	J04177	299	COL11A1, collagen, type XI, alpha 1	AAGCACTAGA [T/G] TTTCACAATT	M	T	G	D	E
G705a15	WIAF-13347	J04177	2225	COL11A1, collagen, type XI, alpha 1	GGGAGCCTGG [G/C] CCTCCAGGTC	S	G	C	G	G

G705u16	WIAF-13679	J04177	5493	COL11A1, collagen, type XI, alpha 1	AATTGATCAA[G/A]TACCTATTGT	M	G	A	V	I
G705u17	WIAF-13700	J04177	3484	COL11A1, collagen, type XI, alpha 1	GGAGTTCAAG[G/A]TCCTGTTGGT	M	G	A	G	D
G705u18	WIAF-13709	J04177	5392	COL11A1, collagen, type XI, alpha 1	GAGATGTCCT[A/T]TGACAATAAT	M	A	T	Y	F
G707u1	WIAF-12363	U32169	4996	COL11A2, collagen, type XI, alpha 2	TCCCCTGAGA[C/T]TCCGTGGGGC	M	C	T	L	F
G707u2	WIAF-12374	U32169	3580	COL11A2, collagen, type XI, alpha 2	CAATGGCGCT[G/A]ATGGCCCACA	M	G	A	D	N
G707u3	WIAF-12385	U32169	2059	COL11A2, collagen, type XI, alpha 2	GCCTGGCTCA[G/A]ACGGACCCCC	M	G	A	D	N
G708a1	WIAF-13354	U73778	1885	COL12A1, collagen, type XII, alpha 1	GCCTCTCCTC[C/T]TGCAGAGACC	M	C	T	P	L
G708a2	WIAF-13355	U73778	3630	COL12A1, collagen, type XII, alpha 1	TGTTGGACAA[G/A]AAATGACAAC	M	G	A	E	K
G708a3	WIAF-13356	U73778	3905	COL12A1, collagen, type XII, alpha 1	GCTTGTTGCA[A/T]GCTGTGGCAA	M	A	T	Q	H
G708a4	WIAF-13357	U73778	7051	COL12A1, collagen, type XII, alpha 1	ATTCCACCAG[C/A]CCGGGATGTA	M	C	A	A	D
G708a5	WIAF-13358	U73778	8036	COL12A1, collagen, type XII, alpha 1	AAGAAGTAAA[G/A]ACATTATTTT	S	G	A	K	K
G708a6	WIAF-13364	U73778	1461	COL12A1, collagen, type XII, alpha 1	TGGCTCCTAT[A/T]GCATTGGGAT	M	A	T	S	C
G708a7	WIAF-13365	U73778	2344	COL12A1, collagen, type XII, alpha 1	ATTACTTGGA[C/T]TCAAGCTCCA	M	C	T	T	I
G708a8	WIAF-13366	U73778	5207	COL12A1, collagen, type XII, alpha 1	CAGATAAGAT[G/A]GAGACCATCT	M	G	A	M	I
G708a9	WIAF-13367	U73778	6592	COL12A1, collagen, type XII, alpha 1	GAGCCCATGG[A/T]AGCCTTTGTT	M	A	T	E	V
G708a10	WIAF-13368	U73778	7434	COL12A1, collagen, type XII, alpha 1	CCAGGATGAG[G/A]TCAAGAAGGC	M	G	A	V	I
G708a11	WIAF-13369	U73778	9108	COL12A1, collagen, type XII, alpha 1	ACCTCGGGGG[C/G]TGCCTGGGCC	M	C	G	L	V
G708a12	WIAF-13370	U73778	9111	COL12A1, collagen, type XII, alpha 1	TCGGGGGCTG[C/T]CTGGGCCCCC	M	C	T	P	S
G708a13	WIAF-13371	U73778	9196	COL12A1, collagen, type XII, alpha 1	CCCCCTGGCC[G/A]TCCTGGAAAC	M	G	A	R	H

G708u14	WIAF-13972	U73778	3044	COL12A1, collagen, type XII, alpha 1	CAGTATTTGC[C/A]ACTTACAGCA	S	C	A	A	A	A	
G708u15	WIAF-13977	U73778	5853	COL12A1, collagen, type XII, alpha 1	TGTGACTGTA[G/C]TTCCCGTTTA	M	G	C	V	L		
G710u1	WIAF-12371	D38163	3082	COL19A1, collagen, type XIX, alpha 1	AGGAAACAAG[G/T]GCTCCATGGG	M	G	T	G	C		
G710u2	WIAF-12388	D38163	2089	COL19A1, collagen, type XIX, alpha 1	TCCAGGGACT[C/T]CAGGGAATGA	M	C	T	P	S		
G711u1	WIAF-12360	L25286	1449	COL15A1, collagen, type XV, alpha 1	TGTGGGTCCA[A/G]GCAGTGAAGA	M	A	G	S	G		
G711u2	WIAF-12372	L25286	4001	COL15A1, collagen, type XV, alpha 1	ATATTCCAAT[A/G]TACTCCTTTG	M	A	G	I	M		
G711u3	WIAF-12373	L25286	3867	COL15A1, collagen, type XV, alpha 1	CCATTTGCAA[G/T]ATCTGTCCAC	M	G	T	D	Y		
G711a4	WIAF-13372	L25286	395	COL15A1, collagen, type XV, alpha 1	CCAGCAGCAC[C/T]CGTGGTGGCG	S	C	T	T	T		
G711a5	WIAF-13373	L25286	3101	COL15A1, collagen, type XV, alpha 1	AAGGCGACCA[G/A]GGAGCCCAGG	S	G	A	Q	Q		
G712u1	WIAF-13619	M92642	3608	COL16A1, collagen, type XVI, alpha 1	GGCGACCAGG[G/A]ATTTCAAGGC	M	G	A	G	E		
G712u2	WIAF-13620	M92642	4944	COL16A1, collagen, type XVI, alpha 1	CCATGAAAAC[C/T]ATGAAGGGGC	S	C	T	T	T		
G712u3	WIAF-13621	M92642	4707	COL16A1, collagen, type XVI, alpha 1	CCAAAGGTGA[A/C]AAAGGGGACA	M	A	C	E	D		
G712u4	WIAF-13654	M92642	421	COL16A1, collagen, type XVI, alpha 1	GCCCACGCGA[C/A]GAGTATTCCC	S	C	A	R	R		
G712u5	WIAF-13655	M92642	444	COL16A1, collagen, type XVI, alpha 1	GGGGTCTCCC[G/A]GAGGAGTTTG	S	G	A	P	P		
G712u6	WIAF-13656	M92642	338	COL16A1, collagen, type XVI, alpha 1	CTCATGAAGA[A/C]GTCTGCCATC	M	A	C	K	T		
G712u7	WIAF-13862	M92642	3227	COL16A1, collagen, type XVI, alpha 1	CCTGGTCCTC[C/T]GGGATTGCCA	M	C	T	P	L		
G712u8	WIAF-13863	M92642	3199	COL16A1, collagen, type XVI, alpha 1	TCCTGGCTGT[G/T]TTGGGAGCCC	M	G	T	V	F		
G712u9	WIAF-13878	M92642	318	COL16A1, collagen, type XVI, alpha 1	ACCTCATCCA[C/T]CGACTCAGCC	S	C	T	H	H		
G712u10	WIAF-13882	M92642	1346	COL16A1, collagen, type XVI, alpha 1	ACAGGCGAGA[A/G]GGGCCAGAAA	M	A	G	K	R		

G712u11	WIAF-13883	M92642	1309	COL16A1, collagen, type XVI, alpha 1	GTCAGGAGCT[C/T]TGGGACCCTC	S	C	T	L	L
G715a1	WIAF-13344	Z74615	3504	COL1A1, collagen, type I, alpha 1	TCCTGGTGAA[C/G]AAGGTCCCTC	M	C	G	Q	E
G717u1	WIAF-12639	Z74616	3988	COL1A2, collagen, type I, alpha 2	ATGAGGAGAC[T/C]GGCAACCTGA	S	T	C	T	T
G720u1	WIAF-12367	X14420	3494	COL3A1, collagen, type III, alpha 1 (Ehlers-Danlos syndrome type IV, autosomal dominant)	GGTGCAATCG[G/A]CAGTCCAGGA	M	G	A	G	D
G720u2	WIAF-12383	X14420	3035	COL3A1, collagen, type III, alpha 1 (Ehlers-Danlos syndrome type IV, autosomal dominant)	GGTGTCAAGG[G/A]TGAAAGTGGG	M	G	A	G	D
G720a3	WIAF-13374	X14420	214	COL3A1, collagen, type III, alpha 1 (Ehlers-Danlos syndrome type IV, autosomal dominant)	TCTTGGTCAG[T/C]CCTATGCGGA	M	T	C	S	P
G720a4	WIAF-13375	X14420	1953	COL3A1, collagen, type III, alpha 1 (Ehlers-Danlos syndrome type IV, autosomal dominant)	CTGGACCTCA[A/G]GGACCCCCAG	S	A	G	Q	Q
G720a5	WIAF-13376	X14420	2194	COL3A1, collagen, type III, alpha 1 (Ehlers-Danlos syndrome type IV, autosomal dominant)	TAGAGGTGGA[G/A]CTGGTCCCCC	M	G	A	A	T
G720a6	WIAF-13377	X14420	3731	COL3A1, collagen, type III, alpha 1 (Ehlers-Danlos syndrome type IV, autosomal dominant)	GGGATTGGAG[G/A]TGAAAAAGCT	M	G	A	G	D
G722u1	WIAF-14132	HT3162	140	COL4A2, collagen, type IV, alpha 2	GAGATTGGCG[C/T]GACTGGTGAT	M	C	T	A	V
G724a1	WIAF-12120	X81053	3892	COL4A4, collagen, type IV, alpha 4	CTCGTGGAAA[G/A]AAAGGTCCCC	S	G	A	K	K
G724a2	WIAF-12121	X81053	4187	COL4A4, collagen, type IV, alpha 4	GAAAGGACCA[A/G]TGGGATTCCC	M	A	G	M	V
G724a3	WIAF-12122	X81053	3802	COL4A4, collagen, type IV, alpha 4	ATGATGTGGG[G/A]CCACCTGGTC	S	G	A	G	G

G724a4	WIAF-12123	X81053	1838	COL4A4, collagen, type IV, alpha 4	ACCAGGAAAG[C/A]ATGGTGCCTC	M	C	A	H	N
G724u5	WIAF-12364	X81053	376	COL4A4, collagen, type IV, alpha 4	CTGTTTGCCA[C/T]TGTGTTCCTG	S	C	T	H	H
G724u6	WIAF-12365	X81053	2018	COL4A4, collagen, type IV, alpha 4	TCCAGGGGAT[C/G]ATGAAGATGC	M	C	G	H	D
G724u7	WIAF-12366	X81053	4756	COL4A4, collagen, type IV, alpha 4	GCCTTCCCGT[A/G]TTTAGCACGC	S	A	G	V	V
G724u8	WIAF-12377	X81053	3595	COL4A4, collagen, type IV, alpha 4	CTGGACCACC[A/G]GGGTGCCCAG	S	A	G	P	P
G724u9	WIAF-12378	X81053	3516	COL4A4, collagen, type IV, alpha 4	GGAGCATCCG[G/C]AGAGCAGGGC	M	G	C	G	A
G724u10	WIAF-12379	X81053	4288	COL4A4, collagen, type IV, alpha 4	CTGGTCTTCC[A/G]GGTCCCAGAG	S	A	G	P	P
G724u11	WIAF-12380	X81053	5140	COL4A4, collagen, type IV, alpha 4	GCCACTTTTT[C/A]GCAAATAAGT	M	C	A	F	L
G724u12	WIAF-12387	X81053	207	COL4A4, collagen, type IV, alpha 4	GACTTGCCTG[C/T]GATGTGGTCT	-	C	T	-	-
G727u1	WIAF-12362	D90279	5135	COL5A1, collagen, type V, alpha 1	TTCAAGGTTT[A/T]CTGCAACTTC	M	A	T	Y	F
G727u2	WIAF-12369	D90279	4686	COL5A1, collagen, type V, alpha 1	AACAGGGTAT[C/T]ACTGGTCCTT	S	C	T	I	I
G727u3	WIAF-12370	D90279	4608	COL5A1, collagen, type V, alpha 1	TCGGTCCTCC[G/C]GGTGAACAGG	S	G	C	P	P
G727a4	WIAF-13300	D90279	2034	COL5A1, collagen, type V, alpha 1	ACGGCCTGGC[T/A]GGGTTGCCAG	S	T	A	A	A
G727a5	WIAF-13301	D90279	2073	COL5A1, collagen, type V, alpha 1	GTGACCCTGG[T/C]CCTTCCGGCC	S	T	C	G	G
G727a6	WIAF-13302	D90279	3763	COL5A1, collagen, type V, alpha 1	CGGGCAGAAA[G/A]GTGATGAAGG	M	G	A	G	S
G729u1	WIAF-11844	L02870	2345	COL7A1, collagen, type VII, alpha 1 (epidermolysis bullosa, dystrophic, dominant and recessive)	ATGGACTGGA[G/A]CCAGATACTG	S	G	A	E	E

G729u2	WIAF-11845	L02870	3083	COL7A1, collagen, type VII, alpha 1 (epidermolysis bullosa, dystrophic, dominant and recessive)	TATCCTGGCG [G/A] CCACTCAGAG	S	G	A	R	R
G729u3	WIAF-11846	L02870	3031	COL7A1, collagen, type VII, alpha 1 (epidermolysis bullosa, dystrophic, dominant and recessive)	GACTCGGTGA [C/T] TTTGGCCTGG	M	C	T	T	I
G729u4	WIAF-11851	L02870	1289	COL7A1, collagen, type VII, alpha 1 (epidermolysis bullosa, dystrophic, dominant and recessive)	CGGACTATGA [G/T] GTGACCGTGA	M	G	T	E	D
G729u5	WIAF-11852	L02870	1032	COL7A1, collagen, type VII, alpha 1 (epidermolysis bullosa, dystrophic, dominant and recessive)	CCAAGTGACT [G/T] TGATTGCCCT	M	G	T	V	L
G729u6	WIAF-11853	L02870	1897	COL7A1, collagen, type VII, alpha 1 (epidermolysis bullosa, dystrophic, dominant and recessive)	CGCCGGGAGC [C/T] GGAAACTCCA	M	C	T	P	L
G729u7	WIAF-11854	L02870	1827	COL7A1, collagen, type VII, alpha 1 (epidermolysis bullosa, dystrophic, dominant and recessive)	GCTTAGCTAC [A/T] CTGTGCGGGT	M	A	T	T	S
G729u8	WIAF-11855	L02870	1893	COL7A1, collagen, type VII, alpha 1 (epidermolysis bullosa, dystrophic, dominant and recessive)	TGTCCGCCGG [G/A] AGCCGGAAAC	M	G	A	E	K

G729u9	WIAF-11864	L02870	2142	COL7A1, collagen, type VII, alpha 1 (epidermolysis bullosa, dystrophic, dominant and recessive)	GGGCCCTGCT [G/A] CAGTCATCGT	M	G	A	A	T
G729u10	WIAF-11865	L02870	2353	COL7A1, collagen, type VII, alpha 1 (epidermolysis bullosa, dystrophic, dominant and recessive)	GAGCCAGATA [C/T] TGAGTATACG	M	C	T	T	I
G729u11	WIAF-11866	L02870	2221	COL7A1, collagen, type VII, alpha 1 (epidermolysis bullosa, dystrophic, dominant and recessive)	TCATCTGTCA [C/T] CATTACCTGG	M	C	T	T	I
G729u12	WIAF-11869	L02870	6585	COL7A1, collagen, type VII, alpha 1 (epidermolysis bullosa, dystrophic, dominant and recessive)	ACCAGGAGAG [C/T] GTGGTATGGC	M	C	T	R	C
G729u13	WIAF-11870	L02870	8169	COL7A1, collagen, type VII, alpha 1 (epidermolysis bullosa, dystrophic, dominant and recessive)	GGGTGACCGA [G/T] GCTTTGACGG	M	G	T	G	C
G729u14	WIAF-11877	L02870	438	COL7A1, collagen, type VII, alpha 1 (epidermolysis bullosa, dystrophic, dominant and recessive)	CGCCATCCGT [G/A] AGCTTAGCTA	M	G	A	E	K
G729u15	WIAF-11882	L02870	3481	COL7A1, collagen, type VII, alpha 1 (epidermolysis bullosa, dystrophic, dominant and recessive)	AGGATCCGTG [A/T] CATGCCCTAC	M	A	T	D	V

G729u16	WIAF-11883	L02870	5654	COL7A1, collagen, type VII, alpha 1 (epidermolysis bullosa, dystrophic, dominant and recessive)	ACGGAGAACC[T/C]GGGGACCCTG	S	T	C	P	P				
G729u17	WIAF-11884	L02870	7124	COL7A1, collagen, type VII, alpha 1 (epidermolysis bullosa, dystrophic, dominant and recessive)	TGCCAGGGCC[G/C]CGAGGCGAGA	S	G	C	P	P				
G729u18	WIAF-11885	L02870	7757	COL7A1, collagen, type VII, alpha 1 (epidermolysis bullosa, dystrophic, dominant and recessive)	GCTTGGATGG[T/C]GACAAAGGAC	S	T	C	G	G				
G729u19	WIAF-13389	L02870	1615	COL7A1, collagen, type VII, alpha 1 (epidermolysis bullosa, dystrophic, dominant and recessive)	ACCGTGGTTC[C/T]CACTGGACCA	M	C	T	P	L				
G729u20	WIAF-13390	L02870	2930	COL7A1, collagen, type VII, alpha 1 (epidermolysis bullosa, dystrophic, dominant and recessive)	TCCTAGGGCC[G/A]GCTGGAGAAG	S	G	A	P	P				
G729u21	WIAF-13399	L02870	5145	COL7A1, collagen, type VII, alpha 1 (epidermolysis bullosa, dystrophic, dominant and recessive)	CCAGGGAGAT[C/T]CTGGAGAGGA	M	C	T	P	S				
G729u22	WIAF-13411	L02870	3472	COL7A1, collagen, type VII, alpha 1 (epidermolysis bullosa, dystrophic, dominant and recessive)	ATCTTGCAAA[G/A]GATCCGTGAC	M	G	A	R	K				
G730a1	WIAF-13303	X57527	305	COL8A1, collagen, type VIII, alpha 1	ATGGGCAAGG[A/G]AGCCGTTCCC	M	A	G	E	G				
G732u1	WIAF-12616	M95610	936	COL9A2, collagen, type IX, alpha 2	CAGGCGGGAC[A/G]GCCCGGAAGT	S	A	G	T	T				

G732u2	WIAF-12617	M95610	696	COL9A2, collagen, type IX, alpha 2	AAGGGAGAGA[C/T]GGGCCCTCAT	S	C	T	D	D
G732u3	WIAF-12619	M95610	1288	COL9A2, collagen, type IX, alpha 2	AAGTGGGTGA[C/T]CCAGGGGTGG	M	C	T	P	S
G732u4	WIAF-12620	M95610	962	COL9A2, collagen, type IX, alpha 2	CCACCAGGGC[C/G]TAGCGGGTGT	M	C	G	P	R
G737u1	WIAF-13394	M13436	?	INHBA, inhibin, beta A (activin A, activin AB alpha polypeptide)	TGCTCCCTG[G/T]	?	G	T		
G738a1	WIAF-13383	M58549	183	MGP, matrix Gla protein	ATGGAGAGCT[A/G]AAGTCCAAGA	M	A	G	K	E
G738a2	WIAF-13384	M58549	330	MGP, matrix Gla protein	GCGCCGAGGG[A/G]CCAAATGAGA	M	A	G	T	A
G739u1	WIAF-11867	U94332	862	TNFRSF11B, tumor necrosis factor receptor superfamily, member 11b (osteoprotegerin)	TGCTGAAGTT[A/G]TGGAAACATC	S	A	G	L	L
G739u2	WIAF-11874	U94332	1244	TNFRSF11B, tumor necrosis factor receptor superfamily, member 11b (osteoprotegerin)	GTATCAGAAG[T/C]TATTTTTAGA	S	T	C	L	L
G743u1	WIAF-13402	HT847	1669	PTHR1, parathyroid hormone receptor 1	CCCTGGAGAC[C/A]CTCGAGACCA	S	C	A	T	T
G747u1	WIAF-12414	J03040	123	SPARC, secreted protein, acidic, cysteine-rich (osteonectin)	CTCAGCAAGA[A/G]GCCCTGCCTG	S	A	G	E	E
G748u1	WIAF-12628	HT0157	117	VDR, vitamin D (1,25-dihydroxyvitamin D3) receptor	CCTTCAGGGA[T/C]GGAGGCAATG	M	T	C	M	T
G748u2	WIAF-12629	HT0157	1171	VDR, vitamin D (1,25-dihydroxyvitamin D3) receptor	CCGCGCTGAT[T/C]GAGGCCATCC	S	T	C	I	I
G748u3	WIAF-12640	HT0157	172	VDR, vitamin D (1,25-dihydroxyvitamin D3) receptor	TTGACCGGAA[C/T]GTGCCCCGGA	S	C	T	N	N
G749u1	WIAF-11862	HT3734	679	osteopontin, alt. transcript 1	ATCACCTCAC[A/T]CATGGAAAGC	M	A	T	H	L
G749u2	WIAF-11875	HT3734	386	osteopontin, alt. transcript 1	AAGATGATGA[A/G]GACCATGTGG	S	A	G	D	D
G749u3	WIAF-11876	HT3734	419	osteopontin, alt. transcript 1	CCATTGACTC[G/A]AACGACTCTG	S	G	A	S	S

G749a4	WIAF-12084	HT3734	171	osteopontin, alt. transcript 1	TAAACAGGCT[G/A]ATTCTGGAAG	M	G	A	D	N
G749u5	WIAF-13387	HT3734	738	osteopontin, alt. transcript 1	CCAGGACCTG[A/C]ACGCGCCTTC	M	A	C	N	H
G749u6	WIAF-13388	HT3734	716	osteopontin, alt. transcript 1	CATACAAGGC[C/A]ATCCCCGTTG	S	C	A	A	A
G751u1	WIAF-12631	HT5036	410	ADM, adrenomedullin	GACAGCAGTC[C/G]GGATGCCGCC	M	C	G	P	R
G752u1	WIAF-11843	HT1782	1405	CHGA, chromogranin A (parathyroid secretory protein 1)	CGGCCATTGA[A/G]GCAGAGCTGG	S	A	G	E	E
G752u2	WIAF-11873	HT1782	1187	CHGA, chromogranin A (parathyroid secretory protein 1)	GGACAACCGG[G/A]ACAGTTCCAT	M	G	A	D	N
G754a1	WIAF-13382	K02043	663	NPPA, natriuretic peptide precursor A	GTACAATGCC[G/A]TGTCCAACGC	M	G	A	V	M
G756u1	WIAF-12395	HT3508	2086	SCNN1A, sodium channel, nonvoltage-gated 1 alpha	CAGTTCCTCC[A/G]CCTGTCCTCT	M	A	G	T	A
G757u1	WIAF-12420	HT28563	797	SCNN1B, sodium channel, nonvoltage-gated 1, beta (Liddle syndrome)	CCTGCAGGCC[A/C]CCAACATCTT	M	A	C	T	P
G757u2	WIAF-12421	HT28563	1006	SCNN1B, sodium channel, nonvoltage-gated 1, beta (Liddle syndrome)	GAACTGAATT[C/T]GGCCTGAAGT	S	C	T	F	F
G757u3	WIAF-12430	HT28563	1768	SCNN1B, sodium channel, nonvoltage-gated 1, beta (Liddle syndrome)	TCATCGACTT[T/C]GTGTGGATCA	S	T	C	F	F
G757u4	WIAF-12494	HT28563	662	SCNN1B, sodium channel, nonvoltage-gated 1, beta (Liddle syndrome)	AAGCAGCTCA[G/C]CATCAGAAAA	M	G	C	A	P
G757u5	WIAF-12506	HT28563	1091	SCNN1B, sodium channel, nonvoltage-gated 1, beta (Liddle syndrome)	GATGCTTCAC[G/C]AGCAGAGGTC	M	G	C	E	Q
G757u6	WIAF-12507	HT28563	1452	SCNN1B, sodium channel, nonvoltage-gated 1, beta (Liddle syndrome)	ACCTGCATTG[G/T]CATGTGCAAG	M	G	T	G	V
G758u1	WIAF-12621	HT27856	415	SCNN1D, sodium channel, nonvoltage-gated 1, delta	CGGGAACCCA[C/T]GTCGGCCGAG	M	C	T	R	C
G758u2	WIAF-12632	HT27856	325	SCNN1D, sodium channel, nonvoltage-gated 1, delta	CCTCTTTGAG[C/T]GTCACTGGCA	M	C	T	R	C

G758u3	WIAF-12634	HT27856	879	SCNN1D, sodium channel, nonvoltage-gated 1, delta	ATGGCGTCTG[G/A]ACAGCTCAGC	N	G	A	W	*
G758u4	WIAF-12635	HT27856	1138	SCNN1D, sodium channel, nonvoltage-gated 1, delta	CGTGGAGGTG[G/C]AGCTGCTACA	M	G	C	E	Q
G762u1	WIAF-12622	HT27531	1850	NPR3, natriuretic peptide receptor C/guanylate cyclase C (atrionatriuretic peptide receptor C)	TAGGAGCTGG[C/T]TTGCTAATGG	S	C	T	G	G
G762u2	WIAF-12623	HT27531	1926	NPR3, natriuretic peptide receptor C/guanylate cyclase C (atrionatriuretic peptide receptor C)	AGAAGAAAGT[A/G]ACCTTGGAAA	M	A	G	N	D
G762u3	WIAF-12624	HT27531	1791	NPR3, natriuretic peptide receptor C/guanylate cyclase C (atrionatriuretic peptide receptor C)	CAAATCATCA[G/T]GTGGCCTAGA	M	G	T	G	C
G762u4	WIAF-12636	HT27531	1963	NPR3, natriuretic peptide receptor C/guanylate cyclase C (atrionatriuretic peptide receptor C)	GAAGATTCCA[T/C]CAGATCCCAT	M	T	C	I	T
G763u1	WIAF-12659	HT3183	1633	NPR2, natriuretic peptide receptor B/guanylate cyclase B (atrionatriuretic peptide receptor B)	CTGGGCCCTT[C/T]CCTGATGAAC	M	C	T	S	F
G763u2	WIAF-12678	HT3183	668	NPR2, natriuretic peptide receptor B/guanylate cyclase B (atrionatriuretic peptide receptor B)	TGCCATCACT[T/C]CTGCTGTTGG	S	T	C	L	L
G763u3	WIAF-12684	HT3183	2354	NPR2, natriuretic peptide receptor B/guanylate cyclase B (atrionatriuretic peptide receptor B)	TGTTTGAACT[C/T]AAACATATGA	S	C	T	L	L

G764u1	WIAF-12698	HT1221	3021	NPR1, natriuretic peptide receptor A/guanylate cyclase A (atrionatriuretic peptide receptor A)	CCCCGTTACT [G/T] TCTCTTTGGG	M	G	T	C	F
G764u2	WIAF-12708	HT1221	588	NPR1, natriuretic peptide receptor A/guanylate cyclase A (atrionatriuretic peptide receptor A)	GAGCGCCAAG [C/T] GCTCATGCTC	M	C	T	A	V
G764u3	WIAF-12709	HT1221	1897	NPR1, natriuretic peptide receptor A/guanylate cyclase A (atrionatriuretic peptide receptor A)	GTCCCCGTGG [G/A] AGCCTGCAGG	S	G	A	G	G
G765u1	WIAF-10012	HT2456	604	DCP1, dipeptidyl carboxypeptidase 1 (angiotensin I converting enzyme)	GCTGGCACAA [A/G] GCTGCGGGCA	S	A	G	N	N
G765u2	WIAF-10014	HT2456	2350	DCP1, dipeptidyl carboxypeptidase 1 (angiotensin I converting enzyme)	TGATGGCCAC [A/G] TCCCGGAAAT	S	A	G	T	T
G765u3	WIAF-10025	HT2456	1688	DCP1, dipeptidyl carboxypeptidase 1 (angiotensin I converting enzyme)	CCCACTGCAC [C/A] AGTGTGACAT	M	C	A	Q	K
G765u4	WIAF-10027	HT2456	3220	DCP1, dipeptidyl carboxypeptidase 1 (angiotensin I converting enzyme)	TCCCCTTCAG [C/T] TACCTCGTCG	S	C	T	S	S
G765u5	WIAF-10028	HT2456	3409	DCP1, dipeptidyl carboxypeptidase 1 (angiotensin I converting enzyme)	TCAGGTACTT [T/C] GTCAGCTTCA	S	T	C	F	F
G765u6	WIAF-10040	HT2456	775	DCP1, dipeptidyl carboxypeptidase 1 (angiotensin I converting enzyme)	AGCCCCTCTA [C/T] CTGAACCTCC	S	C	T	Y	Y

G772u1	WIAF-12626	HT2121	1064	AVPR2, arginine vasopressin receptor 2 (nephrogenic diabetes insipidus)	TCAGCAGCAG[C/T]GTGTCCTCAG	S	C	T	S	S
G772u2	WIAF-12627	HT2121	998	AVPR2, arginine vasopressin receptor 2 (nephrogenic diabetes insipidus)	CCTTTGTGCT[A/G]CTCATGTTGC	S	A	G	L	L
G773u1	WIAF-12644	HT2141	163	SLC6A6, solute carrier family 6 (neurotransmitter transporter, taurine), member 6	CTAGCAAGAT[C/T]GACTTTGTGC	S	C	T	I	I
G773u2	WIAF-12645	HT2141	445	SLC6A6, solute carrier family 6 (neurotransmitter transporter, taurine), member 6	TCGTCATCCT[G/C]GCCTGGGCCA	S	G	C	L	L
G773u3	WIAF-12665	HT2141	289	SLC6A6, solute carrier family 6 (neurotransmitter transporter, taurine), member 6	TGTTTGGGAG[C/T]GGCCTGCCTG	S	C	T	S	S
G773u4	WIAF-12666	HT2141	382	SLC6A6, solute carrier family 6 (neurotransmitter transporter, taurine), member 6	CCTTGTTCTC[T/C]GGTATCGGCT	S	T	C	S	S
G776u1	WIAF-11857	U66088	1457	SLC5A5, solute carrier family 5 (sodium iodide symporter), member 5	TAGAAGACCT[C/T]ATCAAACCTC	S	C	T	L	L
G776u2	WIAF-11871	U66088	2039	SLC5A5, solute carrier family 5 (sodium iodide symporter), member 5	GATTGTTGTG[G/C]TGGGACCTCG	M	G	C	W	C
G776u3	WIAF-13398	U66088	1379	SLC5A5, solute carrier family 5 (sodium iodide symporter), member 5	GGCTTTTCCT[G/A]GCCTGTGCTT	S	G	A	L	L
G777u1	WIAF-12646	HT27843	4348	SMRT	ATACAATATC[A/G]GCCAGCCTGG	M	A	G	S	G
G777u2	WIAF-12654	HT27843	2031	SMRT	CTGAGCTGGG[T/C]AAGCCGCGGC	S	T	C	G	G
G777u3	WIAF-12655	HT27843	2052	SMRT	AGAGCCCCCT[G/A]ACCTATGAGG	S	G	A	L	L
G777u4	WIAF-12675	HT27843	2205	SMRT	CTCGTGAGAT[C/T]GCCAAGTCCC	S	C	T	I	I
G778u1	WIAF-14093	HT1449	8212	TG, thyroglobulin	ATCTCGTCTC[T/C]GAAGACATCT	M	T	C	L	P
G778u2	WIAF-14111	HT1449	6033	TG, thyroglobulin	ATGTGAACGA[C/T]GGTGCGATGC	M	C	T	R	W

G778u3	WIAF-14112	HT1449	6894	TG, thyroglobulin	GTATCTCAAT [G/T] TGTTCATCCC	M	G	T	V	L
G778u4	WIAF-14125	HT1449	2375	TG, thyroglobulin	ATGGGCCTCC [T/C] GAGCAGGTCT	S	T	C	P	P
G778u5	WIAF-14136	HT1449	1931	TG, thyroglobulin	AGGATGTCCA [A/G] TGCTTTTCCG	S	A	G	Q	Q
G783u1	WIAF-12649	X97674	4008	H.sapiens mRNA for transcriptional intermediary factor 2.	CTAGTGGTAT [G/C] CCAGCAACTA	M	G	C	M	I
G783u2	WIAF-12658	X97674	2566	H.sapiens mRNA for transcriptional intermediary factor 2.	GCCTGGCAGT [G/A] AGCTGGACAA	M	G	A	E	K
G783u3	WIAF-12671	X97674	3828	H.sapiens mRNA for transcriptional intermediary factor 2.	CTCTGAGGCC [T/C] GGAGTACCAA	S	T	C	P	P
G785u1	WIAF-13385	HT1291	386	TTR, transthyretin (prealbumin, amyloidosis type I)	CCAACGACTC [C/T] GGCCCCCGCC	S	C	T	S	S
G787u1	WIAF-12652	HT27477	468	TRIP15: thyroid receptor interacting protein 15	GAAAATTATA [T/C] TTAGAACGAG	S	T	C	Y	Y
G792u1	WIAF-12661	HT27476	265	thyroid receptor interactor 14	CAGCTGGAAC [G/A] TGAAGAGGGC	M	G	A	V	M
G793u1	WIAF-12643	HT5152	458	thyroid receptor interactor 8	GGAAGCTTTT [C/G] AAAGAATGTT	N	C	G	S	*
G794u1	WIAF-12664	HT5136	1110	PSMC5, proteasome (prosome, macropain) 26S subunit, ATPase, 5	GCGTGTGCAC [G/A] GAAGCTGGCA	S	G	A	T	T
G797u1	WIAF-11847	HT3919	140	glutamate receptor 3, flip isoform	CTCACGGAGG [A/G] TTCCCCAACA	S	A	G	G	G
G797u2	WIAF-11848	HT3919	759	glutamate receptor 3, flip isoform	GGTTGTGATC [C/T] TAGGGAAACA	S	C	T	L	L
G797u3	WIAF-11849	HT3919	1253	glutamate receptor 3, flip isoform	GCTACTGGAA [C/T] GAGTATGAAA	S	C	T	N	N
G797u4	WIAF-11850	HT3919	1770	glutamate receptor 3, flip isoform	TCTTTTCCTA [G/A] TCAGCAGGTT	M	G	A	V	I
G797u5	WIAF-13404	HT3919	2711	glutamate receptor 3, flip isoform	GCTACAACGT [G/A] TATGGAACAG	S	G	A	V	V
G797u6	WIAF-13405	HT3919	2376	glutamate receptor 3, flip isoform	CTCAGCATTA [G/A] GAACGCCTGT	M	G	A	G	R
G798u1	WIAF-11868	X77748	2655	GRM3, glutamate receptor, metabotropic 3	TGCAGACGAC [A/G] ACCATGTGCA	S	A	G	T	T

G798u2	WIAF-11879	X77748	2771	GRM3, glutamate receptor, metabotropic 3	CACAGACTGC[A/G]CCTCAACAGG	M	A	G	H	R
G798a3	WIAF-12085	X77748	2699	GRM3, glutamate receptor, metabotropic 3	GTGGTCTTGG[G/C]CTGTTTGTTT	M	G	C	G	A
G798a4	WIAF-12086	X77748	2738	GRM3, glutamate receptor, metabotropic 3	ATCCTGTTTC[A/G]ACCCCAGAAG	M	A	G	Q	R
G798a5	WIAF-12087	X77748	2072	GRM3, glutamate receptor, metabotropic 3	ACACCCTTGG[T/C]CAAAGCATCG	M	T	C	V	A
G798a6	WIAF-12088	X77748	2235	GRM3, glutamate receptor, metabotropic 3	CCCTGCTGAC[C/T]AAGACAAACT	S	C	T	T	T
G798u7	WIAF-13391	X77748	1131	GRM3, glutamate receptor, metabotropic 3	GCGCCAATGC[C/T]TCCTTCACCT	S	C	T	A	A
G799u1	WIAF-11880	M81883	2000	GAD1, glutamate decarboxylase 1 (brain, 67kD)	CAACAAATGC[C/T]TGGAACTGGC	S	C	T	L	L
G799u2	WIAF-11881	M81883	1822	GAD1, glutamate decarboxylase 1 (brain, 67kD)	AGGGTATACT[C/T]CAAGGATGCA	S	C	T	L	L
G799u3	WIAF-13392	M81883	661	GAD1, glutamate decarboxylase 1 (brain, 67kD)	GCGTGGCCCA[T/C]GGATGCACCA	S	T	C	H	H
G799u4	WIAF-13393	M81883	556	GAD1, glutamate decarboxylase 1 (brain, 67kD)	AGCTGATGGC[G/A]TCTTCGACCC	S	G	A	A	A
G799u5	WIAF-13410	M81883	1229	GAD1, glutamate decarboxylase 1 (brain, 67kD)	CCTCATGGAA[C/T]AAATAACACT	N	C	T	Q	*
G801u1	WIAF-13403	D49394	1596	HTR3, 5-hydroxytryptamine (serotonin) receptor 3	TTTACCTGCT[A/G]GCGGTGCTGG	S	A	G	L	L
G803a1	WIAF-13118	U66406	1446	EFNB3, ephrin-B3	CTGGGCCTGG[G/A]GGGTGGAGGT	M	G	A	G	E
G804u1	WIAF-11887	Z26653	7237	LAMA2, laminin, alpha 2 (merosin, congenital muscular dystrophy)	TCACTGATGG[G/T]CACATAAAAG	S	G	T	G	G
G804u2	WIAF-11901	Z26653	9351	LAMA2, laminin, alpha 2 (merosin, congenital muscular dystrophy)	GCAAGCCACT[G/C]GAGGTTAATT	M	G	C	W	S
G804u3	WIAF-11924	Z26653	8740	LAMA2, laminin, alpha 2 (merosin, congenital muscular dystrophy)	ACACTACCCG[A/G]AGAATTGGTC	S	A	G	R	R

G804u4	WIAF-11943	Z26653	8577	LAMA2, laminin, alpha 2 (merosin, congenital muscular dystrophy)	ACCAAAATCA[A/G]TGATGGCCAG	M	A	G		N	S			
G804a5	WIAF-12089	Z26653	3372	LAMA2, laminin, alpha 2 (merosin, congenital muscular dystrophy)	CTCTGTGACT[G/A]CTTCCTCCCT	M	G	A	C	Y				
G804a6	WIAF-13227	Z26653	7047	LAMA2, laminin, alpha 2 (merosin, congenital muscular dystrophy)	GTCAGTCCTC[A/g]GGTGGAAGAT	M	A	g	Q	R				
G804u7	WIAF-13437	Z26653	6791	LAMA2, laminin, alpha 2 (merosin, congenital muscular dystrophy)	TGTGAGAGCC[C/T]TGGATGGACC	S	C	T	L	L				
G805u1	WIAF-13416	U14755	799	LHX1, LIM homeobox protein 1	AAGTAACAGC[A/G]GTGTTGCCAA	M	A	G	S	G				
G805u2	WIAF-13417	U14755	743	LHX1, LIM homeobox protein 1	GGCGAGGAAC[T/C]CTACATCATC	M	T	C	L	P				
G805u3	WIAF-13428	U14755	639	LHX1, LIM homeobox protein 1	GCCGTCAGGG[C/A]ATCTCCCCTA	S	C	A	G	G				
G806u1	WIAF-11886	AF026547	2656	CSPG3, chondroitin sulfate proteoglycan 3 (neurocan)	TTGGAGTTCC[A/G]GCCATGTCTA	S	A	G	P	P				
G806u2	WIAF-11895	AF026547	529	CSPG3, chondroitin sulfate proteoglycan 3 (neurocan)	TGACCTTCGC[T/C]GAGGCCCAGG	S	T	C	A	A				
G806u3	WIAF-11896	AF026547	477	CSPG3, chondroitin sulfate proteoglycan 3 (neurocan)	GAGGTGACAG[G/A]TGTTGTGTTC	M	G	A	G	D				
G806u4	WIAF-11917	AF026547	89	CSPG3, chondroitin sulfate proteoglycan 3 (neurocan)	ACAGGATATC[A/G]CCGATGCCAG	M	A	G	T	A				
G806u5	WIAF-11918	AF026547	213	CSPG3, chondroitin sulfate proteoglycan 3 (neurocan)	AGCGCAGCCC[G/C]AGATGCCCCT	M	G	C	R	P				
G806u6	WIAF-11929	AF026547	769	CSPG3, chondroitin sulfate proteoglycan 3 (neurocan)	GCTTTGCCCG[G/A]GAGCTGGGGG	S	G	A	R	R				
G806u7	WIAF-11931	AF026547	3148	CSPG3, chondroitin sulfate proteoglycan 3 (neurocan)	ACATTGATGA[C/T]TGCCTCTGCA	S	C	T	D	D				

G806u8	WIAF-11949	AF026547	209	CSPG3, chondroitin sulfate proteoglycan 3 (neurocan)	GCCAAGCGCA[G/A]CCCGAGATGC	M	G	A	A	T
G806a9	WIAF-13114	AF026547	3430	CSPG3, chondroitin sulfate proteoglycan 3 (neurocan)	ATGAAAACAC[G/A]TGGATCGGCC	S	G	A	T	T
G806u10	WIAF-13420	AF026547	2113	CSPG3, chondroitin sulfate proteoglycan 3 (neurocan)	CCAGGGCAGA[C/G]TTCAGAGAAA	M	C	G	D	E
G806u11	WIAF-13431	AF026547	94	CSPG3, chondroitin sulfate proteoglycan 3 (neurocan)	ATATCACCGA[T/G]GCCAGCGAAA	M	T	G	D	E
G806u12	WIAF-13432	AF026547	275	CSPG3, chondroitin sulfate proteoglycan 3 (neurocan)	ACAGGACTTG[C/T]CCATCCTGGT	M	C	T	P	S
G808a1	WIAF-13117	Y13276	177	TLX, tailless homolog (Drosophila)	GCATGAGCAA[G/a]CCAGCCGGAT	S	G	a	K	K
G810u1	WIAF-11890	X98248	990	SORT1, sortilin 1	ATAAGGATAC[C/A]ACAAGAAGGA	S	C	A	T	T
G810u2	WIAF-11891	X98248	1093	SORT1, sortilin 1	GGCAGCAAAT[G/T]ATGACATGGT	M	G	T	D	Y
G810u3	WIAF-11907	X98248	1683	SORT1, sortilin 1	CAGACGAAGG[T/G]CAATGCTGGC	S	T	G	G	G
G810u4	WIAF-11908	X98248	1433	SORT1, sortilin 1	ATCTCCCAGA[A/C]ACTGAATGTT	M	A	C	K	T
G810u5	WIAF-11909	X98248	1354	SORT1, sortilin 1	GAAGCCTGAA[A/G]ACAGTGAATG	M	A	G	N	D
G810u6	WIAF-11910	X98248	2180	SORT1, sortilin 1	TACCGGAAAA[T/A]TCCAGGGGAC	M	T	A	I	N
G810u7	WIAF-11911	X98248	2264	SORT1, sortilin 1	AACTTTTTGA[G/A]TCCGGAAAAA	M	G	A	S	N
G810u8	WIAF-11925	X98248	1993	SORT1, sortilin 1	TCGAGACTAT[G/A]TTGTGACCAA	M	G	A	V	I
G810u9	WIAF-11939	X98248	1351	SORT1, sortilin 1	GAGGAAGCCT[G/C]AAAACAGTGA	M	G	C	E	Q
G810u10	WIAF-11940	X98248	2232	SORT1, sortilin 1	AAGTAAAAGA[C/T]TTGAAAAAGA	S	C	T	D	D
G810a11	WIAF-13115	X98248	1769	SORT1, sortilin 1	TCCATGAATA[T/A]CAGCATTTGG	M	T	A	I	N
G810a12	WIAF-13116	X98248	1757	SORT1, sortilin 1	CCTGGAGCTA[G/A]GTCCATGAAT	M	G	A	R	K
G811u1	WIAF-11893	HT3676	900	synapsin I, alt. transcript 1	TGACCAAGAC[G/A]TATGCCACTG	S	G	A	T	T
G811u2	WIAF-11894	HT3676	758	synapsin I, alt. transcript 1	ACCTTCTACC[C/T]CAATCACAAA	M	C	T	P	L
G811u3	WIAF-11927	HT3676	996	synapsin I, alt. transcript 1	CGTCAGTGTC[A/T]GGGAACTGGA	S	A	T	S	S
G811u4	WIAF-11928	HT3676	1054	synapsin I, alt. transcript 1	CATGTCTGAC[A/G]GATACAAGCT	M	A	G	R	G
G811u5	WIAF-13418	HT3676	249	synapsin I, alt. transcript 1	TGTCCAACGC[G/A]GTCAAGCAGA	S	G	A	A	A

G811u6	WIAF-13419	HT3676	432	synapsin I, alt. transcript 1	TTAAAGTAGA[G/A]CAGGCCGAAT	S	G	A	E	E
G812u1	WIAF-11898	HT4564	163	STX1A, syntaxin 1A (brain)	CCAACCCCGA[T/C]GAGAAGACGA	S	T	C	D	D
G812u2	WIAF-11942	HT4564	604	STX1A, syntaxin 1A (brain)	TACACGACAT[G/T]TTCATGGACA	M	G	T	M	I
G813u1	WIAF-11934	U72508	939	Human B7 mRNA, complete cds.	TATGACAGAG[G/A]ACAGAGGATG	M	G	A	G	E
G813u2	WIAF-11948	U72508	619	Human B7 mRNA, complete cds.	GCATCCACAT[G/C]GTGACAGGTC	M	G	C	M	I
G816u1	WIAF-11897	HT4230	151	HTR2B, 5-hydroxytryptamine (serotonin) receptor 2B	CTAACTGGTC[T/G]GGATTACAGA	S	T	G	S	S
G816u2	WIAF-11930	HT4230	189	HTR2B, 5-hydroxytryptamine (serotonin) receptor 2B	GAAATGAAAC[A/G]GATTGTTGAG	M	A	G	Q	R
G818u1	WIAF-11902	HT2694	753	TPH, tryptophan hydroxylase (tryptophan 5-monooxygenase)	GAGTTTTTCA[C/T]TGCACTCAAT	S	C	T	H	H
G818u2	WIAF-11903	HT2694	775	TPH, tryptophan hydroxylase (tryptophan 5-monooxygenase)	TGTGAGACAC[A/G]GTTCAGATCC	M	A	G	S	G
G818u3	WIAF-11904	HT2694	1211	TPH, tryptophan hydroxylase (tryptophan 5-monooxygenase)	TATAATCCAT[A/C]TACACGGAGT	M	A	C	Y	S
G818u4	WIAF-11905	HT2694	1081	TPH, tryptophan hydroxylase (tryptophan 5-monooxygenase)	GATTACCTGC[A/C]AACAGGAATG	M	A	C	K	Q
G818u5	WIAF-11933	HT2694	795	TPH, tryptophan hydroxylase (tryptophan 5-monooxygenase)	CCTTCTATAC[C/T]CCAGAGCCAG	S	C	T	T	T
G818u6	WIAF-11935	HT2694	1239	TPH, tryptophan hydroxylase (tryptophan 5-monooxygenase)	TCCTGAAAGA[C/T]ACCAAGAGCA	S	C	T	D	D
G822u1	WIAF-11906	HT0207	936	ASMT, acetylserotonin N-methyltransferase	CAGACGGAAA[G/T]TGCTCACACC	M	G	T	K	N
G822u2	WIAF-11919	HT0207	637	ASMT, acetylserotonin N-methyltransferase	TGGTGGGACA[C/T]GGATAAAGCT	M	C	T	R	W

G822u3	WIAF-11936	HT0207	318	ASMT, acetylserotonin N-methyltransferase	GAAAAGCTTT[C/T]TATCGAAACA	S	C	T	F	F
G822u4	WIAF-11937	HT0207	116	ASMT, acetylserotonin N-methyltransferase	AATGACTACG[C/T]CAACGGCTTC	M	C	T	A	V
G822u5	WIAF-11938	HT0207	930	ASMT, acetylserotonin N-methyltransferase	ACTGGGCAGA[C/T]GGAAAGTGCT	S	C	T	D	D
G822u6	WIAF-13427	HT0207	120	ASMT, acetylserotonin N-methyltransferase	ACTACGCCAA[C/A]GGCTTCATGG	M	C	A	N	K
G825u1	WIAF-11888	HT4974	236	ADAR, adenosine deaminase, RNA-specific	GCTCAGATAC[C/T]AGCAGCCTGG	N	C	T	Q	*
G825u2	WIAF-11900	HT4974	3076	ADAR, adenosine deaminase, RNA-specific	TCTTTGACAA[A/G]TCCTGCAGCG	S	A	G	K	K
G825u3	WIAF-11912	HT4974	2537	ADAR, adenosine deaminase, RNA-specific	CTTGATTGGG[G/C]AGAACGAGAA	M	G	C	E	Q
G825u4	WIAF-11941	HT4974	3558	ADAR, adenosine deaminase, RNA-specific	GATGGCTATG[A/G]CCTGGAGATC	M	A	G	D	G
G825a5	WIAF-12090	HT4974	1305	ADAR, adenosine deaminase, RNA-specific	CCTGAGACCA[A/G]AAGAAACGCA	M	A	G	K	R
G825u6	WIAF-13426	HT4974	3683	ADAR, adenosine deaminase, RNA-specific	CCGCAGGGAT[C/T]TACTGAGACT	S	C	T	L	L
G826u1	WIAF-12554	X99383	2109	ADARB1, adenosine deaminase, RNA-specific, B1 (homolog of rat RED1)	AGATTACCAA[A/G]CCCAACGTGT	S	A	G	K	K
G826u2	WIAF-12566	X99383	1698	ADARB1, adenosine deaminase, RNA-specific, B1 (homolog of rat RED1)	TGTCCTGCAG[T/G]GACAAGATTG	M	T	G	S	R
G829u1	WIAF-13735	U49262	1404	DVL3, dishevelled 3 (homologous to Drosophila dsh)	GGGTTGGAGG[T/C]CCGTGACTGC	M	T	C	V	A
G83u1	WIAF-10449	HT1576	1338	DNMT1, DNA (cytosine-5-)-methyltransferase 1	ATGATGACCC[G/A]TCTCTTGAAG	S	G	A	P	P
G83u2	WIAF-10450	HT1576	1871	DNMT1, DNA (cytosine-5-)-methyltransferase 1	AAGCTGGTCT[A/G]CCAGATCTTC	M	A	G	Y	C
G83u3	WIAF-10468	HT1576	928	DNMT1, DNA (cytosine-5-)-methyltransferase 1	AAATCCACAG[A/G]TTTCTGATGA	M	A	G	I	V
G83u4	WIAF-10469	HT1576	1562	DNMT1, DNA (cytosine-5-)-methyltransferase 1	AATTCCGACT[C/T]GACCTATGAG	M	C	T	S	L
G83u5	WIAF-10471	HT1576	2424	DNMT1, DNA (cytosine-5-)-methyltransferase 1	GGGCCACGTC[G/A]GACCCTCTGG	S	G	A	S	S

G83u6	WIAF-10473	HT1576	3790	DNMT1, DNA (cytosine-5-)-methyltransferase 1	GTTCTTCCTC[C/T]TGGAGAATGT	S	C	T	L	L
G83u7	WIAF-10486	HT1576	1581	DNMT1, DNA (cytosine-5-)-methyltransferase 1	AGGACCTGAT[C/A]AACAAGATCG	S	C	A	I	I
G832u1	WIAF-12577	L13387	1129	PAFAH1B1, platelet-activating factor acetylhydrolase, isoform Ib, alpha subunit (45kD)	AGACATTCAC[A/T]GGACACAGAG	S	A	T	T	T
G835u1	WIAF-12555	U38276	1311	SEMA3F, sema domain, immunoglobulin domain (Ig), short basic domain, secreted, 3F	CCTCTGGCTC[C/A]GTGTTCCGAG	S	C	A	S	S
G835u2	WIAF-12556	U38276	1229	SEMA3F, sema domain, immunoglobulin domain (Ig), short basic domain, secreted, 3F	ACTCACTTTG[A/T]TGAGCTCCAG	M	A	T	D	V
G835u3	WIAF-12557	U38276	1473	SEMA3F, sema domain, immunoglobulin domain (Ig), short basic domain, secreted, 3F	GAACCTTCAC[G/A]CCATCTATGA	S	G	A	T	T
G835a4	WIAF-13138	U38276	1726	SEMA3F, sema domain, immunoglobulin domain (Ig), short basic domain, secreted, 3F	TGACCAGGAG[A/T]TGGAGGAGCT	M	A	T	M	L
G836u1	WIAF-12592	U28369	1056	SEMA3B, sema domain, immunoglobulin domain (Ig), short basic domain, secreted, 3B	AACGACGTGG[G/A]CGGCCAGCGC	M	G	A	G	D
G836u2	WIAF-12609	U28369	1479	SEMA3B, sema domain, immunoglobulin domain (Ig), short basic domain, secreted, 3B	GTCCTGCCCA[C/T]TGGGGGGCGC	M	C	T	T	I
G838u1	WIAF-12590	U72671	1107	ICAM5, intercellular adhesion molecule 5, telencephalin	CGCAGCTGGG[A/G]CCCAAGCTCT	M	A	G	T	A
G838u2	WIAF-12591	U72671	966	ICAM5, intercellular adhesion molecule 5, telencephalin	CAGGCAGCTG[A/G]TCTGCAACGT	M	A	G	I	V

G840a1	WIAF-12109	HT961	2232	SOS1, son of sevenless (Drosophila) homolog 1	CTCAGGCAAA[T/C]GGAGTAAGCC	S	T	C	N	N	
G840a2	WIAF-12110	HT961	2404	SOS1, son of sevenless (Drosophila) homolog 1	ACCGTCTGAA[C/G]TTGTAGGGAG	M	C	G	L	V	
G840u3	WIAF-12213	HT961	3813	SOS1, son of sevenless (Drosophila) homolog 1	CAAGGGTACC[G/A]CGTCGATGCT	S	G	A	P	P	
G841u1	WIAF-12153	HT97420	1372	SMOH, smoothened (Drosophila) homolog	TTTTGGCTTC[C/G]TGGCCTTTGG	M	C	G	L	V	
G841u2	WIAF-12179	HT97420	858	SMOH, smoothened (Drosophila) homolog	CCCAGTTCAT[G/T]GATGGTGCCC	M	G	T	M	I	
G841u3	WIAF-12185	HT97420	1164	SMOH, smoothened (Drosophila) homolog	CTGTGAGTGG[C/G]ATTTGTTTTG	S	C	G	G	G	
G847u1	WIAF-12588	L41939	2019	EPHB2, EphB2	GGTCTGCAGT[G/T]GCCACCTGAA	M	G	T	G	C	
G847u2	WIAF-12596	L41939	1806	EPHB2, EphB2	GTGTAACAGA[A/C]GACGGGGGTT	S	A	C	R	R	
G847u3	WIAF-12613	L41939	2885	EPHB2, EphB2	AGGCCATCAA[G/C]ATGGGGCAGT	M	G	C	K	N	
G848u1	WIAF-12685	L40636	2484	EPHB1, EphB1	GTCAACAGTA[A/G]CCTGGTGTGC	M	A	G	N	S	
G848u2	WIAF-12690	L40636	2020	EPHB1, EphB1	CCTTCACTTA[T/C]GAGGATCCCA	S	T	C	Y	Y	
G849u1	WIAF-11920	D83492	1544	EPHB6, EphB6	ACCTGTGTGG[C/T]TCATGCAGAG	M	C	T	A	V	
G849u2	WIAF-11921	D83492	3301	EPHB6, EphB6	CTTTGGGATA[C/T]TCATGTGGGA	M	C	T	L	F	
G849u3	WIAF-13412	D83492	1139	EPHB6, EphB6	GAGACCTTCA[C/T]CCTTTACTAC	M	C	T	T	I	
G849u4	WIAF-13413	D83492	1895	EPHB6, EphB6	TTTGAGGTGC[A/C]AGGCTCAGCA	M	A	C	Q	P	
G849u5	WIAF-13414	D83492	2338	EPHB6, EphB6	CTATGACCAG[G/A]CAGAAGACGA	M	G	A	A	T	
G849u6	WIAF-13415	D83492	2567	EPHB6, EphB6	GGGGCTTTGG[C/G]CTTCCTCCTG	M	C	G	A	G	
G849u7	WIAF-13422	D83492	2860	EPHB6, EphB6	GGCCATCCAG[G/A]CCCTGTGGGC	M	G	A	A	T	
G849u8	WIAF-13423	D83492	2782	EPHB6, EphB6	GGAGGTCATT[G/C]GGACAGGCTC	M	G	C	G	R	
G849u9	WIAF-13424	D83492	3038	EPHB6, EphB6	TTCCTCAGGC[A/G]GCGGCAGGGC	M	A	G	Q	R	
G849u10	WIAF-13425	D83492	3637	EPHB6, EphB6	AGCCATTGGA[C/T]TGGAGTGCTA	S	C	T	L	L	
G856u1	WIAF-12625	D45906	1323	LIMK2, LIM domain kinase 2	AGCTGAACCT[G/C]CTGACAGAGT	S	G	C	L	L	
G858u1	WIAF-12630	U65019	864	MADH2, MAD (mothers against decapentaplegic, Drosophila) homolog 2	TTTGGTGTTC[G/A]ATAGCATATT	S	G	A	S	S	
G86u1	WIAF-10437	HT1701	263	RAD51, RAD51 (S. cerevisiae) homolog (E coli RecA homolog)	TGAAGCAAAT[G/C]CAGATACTTC	M	G	C	A	P	
G86u2	WIAF-10465	HT1701	861	RAD51, RAD51 (S. cerevisiae) homolog (E coli RecA homolog)	GCATCAGCCA[T/C]GATGGTAGAA	M	T	C	M	T	

G86u3	WIAF-10466	HT1701	924	RAD51, RAD51 (S. cerevisiae) homolog (E coli RecA homolog)	TACAGAACAG[A/G]CTACTCGGGT	M	A	G	D	G
G864a1	WIAF-13139	X82324	183	POU3F4, POU domain, class 3, transcription factor 4	CAGCAATGGG[C/t]ATCCCCTCGG	M	C	t	H	Y
G866u1	WIAF-12637	HT0101	2576	glutamate receptor (GB:M64752)	AAATCCCGTA[G/A]TGAATCCAAG	M	G	A	S	N
G866u2	WIAF-12638	HT0101	1131	glutamate receptor (GB:M64752)	TAACAGGAAA[C/T]GTGCAGTTTA	S	C	T	N	N
G869u1	WIAF-13406	HT33620	3627	GRIN2C, glutamate receptor, ionotropic, N-methyl D-aspartate 2C	AGATCAGCAG[G/T]GTAGCCCGTG	M	G	T	R	S
G870u1	WIAF-11889	HT4468	714	SLC1A1, solute carrier family 1 (neuronal/epithelial high affinity glutamate transporter, system Xag), member 1	CAGAAGAGTC[C/G]TTCACAGCTG	S	C	G	S	S
G870u2	WIAF-11913	HT4468	314	SLC1A1, solute carrier family 1 (neuronal/epithelial high affinity glutamate transporter, system Xag), member 1	CTAGAGAAAT[T/A]CTACTTTGCT	M	T	A	F	Y
G870u3	WIAF-11914	HT4468	579	SLC1A1, solute carrier family 1 (neuronal/epithelial high affinity glutamate transporter, system Xag), member 1	AAGTCAGTAC[G/A]GTGGATGCCA	S	G	A	T	T
G870u4	WIAF-11922	HT4468	706	SLC1A1, solute carrier family 1 (neuronal/epithelial high affinity glutamate transporter, system Xag), member 1	GAACATGACA[G/A]AAGAGTCCTT	M	G	A	E	K

G870u5	WIAF-11923	HT4468	978	SLC1A1, solute carrier family 1 (neuronal/epithelial high affinity glutamate transporter, system Xag), member 1	GGAAGATCAT [A/G] GAAGTTGAAG	M	A	G	I	M
G871u1	WIAF-11892	HT3187	1004	SLC1A3, solute carrier family 1 (glial high affinity glutamate transporter), member 3	TTCTCTTAAC [G/C] AAGCCATCAT	M	G	C	E	Q
G871u2	WIAF-11915	HT3187	1154	SLC1A3, solute carrier family 1 (glial high affinity glutamate transporter), member 3	TGTTGGCTTA [C/T] TCATTCACGC	M	C	T	L	F
G871u3	WIAF-11926	HT3187	1412	SLC1A3, solute carrier family 1 (glial high affinity glutamate transporter), member 3	GGCTGCCATT [T/G] TCATTGCTCA	M	T	G	F	V
G871u4	WIAF-11944	HT3187	1217	SLC1A3, solute carrier family 1 (glial high affinity glutamate transporter), member 3	AAACCCTTGG [G/A] TTTTTATTGG	M	G	A	V	I
G872u1	WIAF-13433	HT4077	1271	SLC1A2, solute carrier family 1 (glial high affinity glutamate transporter), member 2	CTGTTGGAGC [A/C] ACCATTAACA	S	A	C	A	A
G879u1	WIAF-11899	HT28317	1273	GRM2, glutamate receptor, metabotropic 2	GACTTTGTGC [T/C] CAACGTCAAG	M	T	C	L	P
G879u2	WIAF-11932	HT28317	2349	GRM2, glutamate receptor, metabotropic 2	CTTCTATGTC [A/G] CCTCCAGTGA	M	A	G	T	A
G879u3	WIAF-13421	HT28317	2186	GRM2, glutamate receptor, metabotropic 2	ATGCAAGTAT [G/T] TTGGGCTCGC	M	G	T	M	I
G879u4	WIAF-13429	HT28317	2567	GRM2, glutamate receptor, metabotropic 2	CCCAGTTTGT [C/T] CCCACTGTTT	S	C	T	V	V
G879u5	WIAF-13436	HT28317	2046	GRM2, glutamate receptor, metabotropic 2	ACAGGTGGCC [A/G] TCTGCCTGGC	M	A	G	I	V
G879u6	WIAF-13438	HT28317	2425	GRM2, glutamate receptor, metabotropic 2	GTGCTTGGCT [G/T] CCTCTTTGCG	M	G	T	C	F

G879u7	WIAF-13439	HT28317	2463	GRM2, glutamate receptor, metabotropic 2	CCTCTTCCAG[C/T]CGCAGAAGAA	M	C	T	P	S
G880u1	WIAF-12164	HT33719	2117	GRM4, glutamate receptor, metabotropic 4	AGCCCGACCT[T/G]GGCACCTGCT	S	T	G	L	L
G880u2	WIAF-12176	HT33719	2427	GRM4, glutamate receptor, metabotropic 4	GGACCTGTCG[C/T]TCATCTGCCT	M	C	T	L	F
G880u3	WIAF-12192	HT33719	2372	GRM4, glutamate receptor, metabotropic 4	ACCAGCGGAC[A/G]CTCGACCCCC	S	A	G	T	T
G883a1	WIAF-13140	HT48863	1408	GRM7, glutamate receptor, metabotropic 7	ATCGCAAATG[C/a]ACAGGACAGG	N	C	a	C	*
G883a2	WIAF-13141	HT48863	2027	GRM7, glutamate receptor, metabotropic 7	TCCTGTCTTC[C/t]TGGCAATGTT	S	C	t	L	L
G883a3	WIAF-13147	HT48863	1813	GRM7, glutamate receptor, metabotropic 7	TGTGCACACT[A/g]CCATGTAAGC	S	A	g	L	L
G883a4	WIAF-13148	HT48863	1536	GRM7, glutamate receptor, metabotropic 7	TGTGCTGACT[A/t]CCGGGGTGTC	M	A	t	Y	F
G883a5	WIAF-13149	HT48863	2473	GRM7, glutamate receptor, metabotropic 7	AAGCCAGAGG[G/a]GTTCTCAAGT	S	G	a	G	G
G883a6	WIAF-13150	HT48863	2434	GRM7, glutamate receptor, metabotropic 7	TCATAGACTA[C/t]GATGAACACA	S	C	t	Y	Y
G884u1	WIAF-11916	U95025	1052	GRM8, glutamate receptor, metabotropic 8	CGAACTCTTG[C/A]CAATAATCGA	M	C	A	A	D
G884u2	WIAF-11945	U95025	2016	GRM8, glutamate receptor, metabotropic 8	AAACAAACCG[T/C]ATCCACCGAA	S	T	C	R	R
G884u3	WIAF-11946	U95025	1852	GRM8, glutamate receptor, metabotropic 8	GAGGGCTTCA[G/A]GACGCGAACT	M	G	A	G	R
G884u4	WIAF-11947	U95025	2078	GRM8, glutamate receptor, metabotropic 8	ATTAGTCCAG[C/G]ATCTCAGCTG	M	C	G	A	G
G884u5	WIAF-13430	U95025	1897	GRM8, glutamate receptor, metabotropic 8	TTTTCTCTGT[T/G]ATTCAATCAC	M	T	G	Y	D
G884u6	WIAF-13435	U95025	2364	GRM8, glutamate receptor, metabotropic 8	TTACCATGTA[T/C]ACCACCTGCA	S	T	C	Y	Y
G885u1	WIAF-13434	AF002700	1363	GFRA2, GDNF family receptor alpha 2	AACTCAGGCC[C/A]CAGCAGAGCC	M	C	A	P	H
G886a1	WIAF-13142	U95847	497	GFRA1, GDNF family receptor alpha 1	GAAGTCGCTC[T/a]ACAACTGCCG	M	T	a	Y	N
G886a2	WIAF-13143	U95847	1385	GFRA1, GDNF family receptor alpha 1	GTCTGAGAAT[G/a]AAATTCCCAC	M	G	a	E	K

G886a3	WIAF-13151	U95847	781	GFRA1, GDNF family receptor alpha 1	GCGTGTCCAA [T/c] GATGTCTGCA	S	T	C	N	N
G892u1	WIAF-11956	U12140	798	NTRK2, neurotrophic tyrosine kinase, receptor, type 2	TGGGCAATCC [A/G] TTTACATGCT	S	A	G	P	P
G892u2	WIAF-11957	U12140	834	NTRK2, neurotrophic tyrosine kinase, receptor, type 2	GGATCAAGAC [T/A] CTCCAAGAGG	S	T	A	T	T
G892u3	WIAF-11958	U12140	956	NTRK2, neurotrophic tyrosine kinase, receptor, type 2	GCAAATCTGG [C/T] CGCACCTAAC	M	C	T	A	V
G892u4	WIAF-11960	U12140	1738	NTRK2, neurotrophic tyrosine kinase, receptor, type 2	CTCCAAGTTT [G/A] GCATGAAAGG	M	G	A	G	S
G892u5	WIAF-11962	U12140	2486	NTRK2, neurotrophic tyrosine kinase, receptor, type 2	GTCGGTGGCC [A/G] CACAATGCTG	M	A	G	H	R
G892u6	WIAF-11965	U12140	1106	NTRK2, neurotrophic tyrosine kinase, receptor, type 2	TCCTTAAGGA [T/C] AACTAACATT	M	T	C	I	T
G892u7	WIAF-11966	U12140	2085	NTRK2, neurotrophic tyrosine kinase, receptor, type 2	AGGATGCCAG [T/C] GACAATGCAC	S	T	C	S	S
G892u8	WIAF-11967	U12140	2230	NTRK2, neurotrophic tyrosine kinase, receptor, type 2	GGACCTCAAC [A/C] AGTTCCTCAG	M	A	C	K	Q
G892u9	WIAF-11968	U12140	2223	NTRK2, neurotrophic tyrosine kinase, receptor, type 2	AGCATGGGGA [C/T] CTCAACAAGT	S	C	T	D	D
G892u10	WIAF-11992	U12140	1602	NTRK2, neurotrophic tyrosine kinase, receptor, type 2	GTAATGAAAT [C/T] CCTTCCACAG	S	C	T	I	I
G892u11	WIAF-11998	U12140	1354	NTRK2, neurotrophic tyrosine kinase, receptor, type 2	TACTAAAATA [C/T] ATGTTACCAA	M	C	T	H	Y
G892u12	WIAF-11999	U12140	1944	NTRK2, neurotrophic tyrosine kinase, receptor, type 2	CATTTGTTCA [G/C] CACATCAAGC	M	G	C	Q	H

G892u13	WIAF-12000	U12140	2103	NTRK2, neurotrophic tyrosine kinase, receptor, type 2	CACGCAAGGA[C/T]TTCCACCGTG	S	C	T	D												
G892u14	WIAF-12001	U12140	1860	NTRK2, neurotrophic tyrosine kinase, receptor, type 2	CTGTCATTAT[T/C]GGAATGACCA	S	T	C	I	I											
G892a15	WIAF-13144	U12140	1868	NTRK2, neurotrophic tyrosine kinase, receptor, type 2	ATTGGAATGA[C/G]CAAGATCCCT	M	C	G	T	S											
G892a16	WIAF-13145	U12140	1903	NTRK2, neurotrophic tyrosine kinase, receptor, type 2	CCAGTACTTT[G/T]GCATCACCAA	M	G	T	G	C											
G892a17	WIAF-13146	U12140	1965	NTRK2, neurotrophic tyrosine kinase, receptor, type 2	GACATAACAT[T/G]GTTCTGAAAA	M	T	G	I	M											
G892u18	WIAF-13442	U12140	958	NTRK2, neurotrophic tyrosine kinase, receptor, type 2	AAATCTGGCC[G/T]CACCTAACCT	M	G	T	A	S											
G892u19	WIAF-13446	U12140	2502	NTRK2, neurotrophic tyrosine kinase, receptor, type 2	TGCTGCCCAT[T/C]CGCTGGATGC	S	T	C	I	I											
G892u20	WIAF-13447	U12140	2317	NTRK2, neurotrophic tyrosine kinase, receptor, type 2	GATGCTGCAT[A/T]TAGCCCAGCA	M	A	T	I	L											
G892u21	WIAF-13448	U12140	2364	NTRK2, neurotrophic tyrosine kinase, receptor, type 2	CGTCCCAGCA[C/A]TTCGTGCACC	M	C	A	H	Q											
G892u22	WIAF-13449	U12140	2507	NTRK2, neurotrophic tyrosine kinase, receptor, type 2	CCCATTCGCT[G/A]GATGCCTCCA	N	G	A	W	*											
G892u23	WIAF-13471	U12140	2389	NTRK2, neurotrophic tyrosine kinase, receptor, type 2	TTTGGCCACC[A/C]GGAACTGCCT	S	A	C	R	R											
G892u24	WIAF-13472	U12140	2416	NTRK2, neurotrophic tyrosine kinase, receptor, type 2	GGAGAACTTG[C/T]TGGTGAAAAT	S	C	T	L	L											
G892u25	WIAF-13474	U12140	359	NTRK2, neurotrophic tyrosine kinase, receptor, type 2	GGGATGTCGT[C/T]CTGGATAAGG	M	C	T	S	F											

G892u26	WIAF-13479	U12140	1044	NTRK2, neurotrophic tyrosine kinase, receptor, type 2	TGTATTGGGA[T/C]GTTGGTAACC	S	T	C	D	D
G9u1	WIAF-10222	J03826	1130	FDXR, ferredoxin reductase	GGTATAAGAG[C/T]CGCCCTGTCG	S	C	T	S	S
G9u2	WIAF-10258	J03826	388	FDXR, ferredoxin reductase	CCGGAGCTGC[A/G]GGAGGCCTAC	M	A	G	Q	R
G900u1	WIAF-11970	HT3470	497	STX4A, syntaxin 4A (placental)	TGCAATTCAA[T/C]GCAGTCCGAA	M	T	C	M	T
G901u1	WIAF-11969	HT27792	758	STX3A, syntaxin 3A	TGCACACAGT[G/A]GACCACGTGG	S	G	A	V	V
G901u2	WIAF-11971	HT27792	317	STX3A, syntaxin 3A	ACGTCCGGAA[C/A]AAACTGAAGA	M	C	A	N	K
G901u3	WIAF-12002	HT27792	611	STX3A, syntaxin 3A	AGCAAGCCCT[C/T]AGTGAGATTG	S	C	T	L	L
G901u4	WIAF-12003	HT27792	909	STX3A, syntaxin 3A	GCTGAATTAA[G/A]AGTGGCCTAA	-	G	A	-	-
G901u5	WIAF-12004	HT27792	163	STX3A, syntaxin 3A	ATTGAGGAAA[C/T]TCGGCTTAAC	M	C	T	T	I
G901a6	WIAF-13152	HT27792	82	STX3A, syntaxin 3A	CAGCTGACAC[A/G]GGATGATGAT	M	A	G	Q	R
G901u7	WIAF-13453	HT27792	828	STX3A, syntaxin 3A	CCGGAAGAAA[T/C]TGATAATTAT	S	T	C	L	L
G901u8	WIAF-13455	HT27792	226	STX3A, syntaxin 3A	TACAGTATCA[T/C]TCTCTCTGCA	M	T	C	I	T
G902u1	WIAF-13454	HT27744	848	STX5A, syntaxin 5A	ACTTCCAGTC[T/A]GTCACCTCCA	S	T	A	S	S
G902u2	WIAF-13456	HT27744	338	STX5A, syntaxin 5A	ATTTCGTGAG[A/G]GCCAAGGGCA	S	A	G	R	R
G905u1	WIAF-12202	HT27789	487	CREBL1, cAMP responsive element binding protein-like 1	TCCAGATCAA[C/T]GTTATCCCCA	S	C	T	N	N
G905u2	WIAF-12219	HT27789	151	CREBL1, cAMP responsive element binding protein-like 1	ATTCTGGCCT[A/T]GATGAAGTGG	S	A	T	L	L
G905u3	WIAF-12230	HT27789	649	CREBL1, cAMP responsive element binding protein-like 1	AGTCCCTGTC[C/G]CCTTCAGGAT	S	C	G	S	S
G906u1	WIAF-12214	HT4372	2127	N-ethylmaleimide-sensitive factor	AAGGGAAGAA[G/A]GTCTGGATAG	S	G	A	K	K
G906u2	WIAF-12221	HT4372	514	N-ethylmaleimide-sensitive factor	GGGAGAGCCT[G/A]CGACAGGGAA	M	G	A	A	T
G908u1	WIAF-12201	HT3665	98	RAB5A, RAB5A, member RAS oncogene family	GCCCAAATAC[T/G]GGAAATAAAA	S	T	G	T	T

G91u1	WIAF-10438	HT1848	496	ERCC1, excision repair cross-complementing rodent repair deficiency, complementation group 1 (includes overlapping antisense sequence)	TCGTGCGCAA[C/T]GTGCCCTGGG	S	C	T	N	N
G91u2	WIAF-10439	HT1848	367	ERCC1, excision repair cross-complementing rodent repair deficiency, complementation group 1 (includes overlapping antisense sequence)	CTGGGGCCAC[G/A]TGCCCCACAG	S	G	A	T	T
G914a1	WIAF-13210	HT3672	252	synaptobrevin 1	GCAGTGCTGC[C/A]AAGCTAAAGA	S	C	A	A	A
G915a1	WIAF-12115	D63506	1390	Homo sapiens mRNA for unc-18homologue, complete cds.	TTACCTTGGT[G/A]TTCCCATTGT	M	G	A	V	I
G915u2	WIAF-12293	D63506	685	Homo sapiens mRNA for unc-18homologue, complete cds.	ACAGCTTGTT[G/A]AAAAAAAGCT	M	G	A	E	K
G916a1	WIAF-13209	HT28523	308	Huntingtin associated protein 1-like protein	GAGCAGTTTT[C/T]GGAGGCCAGC	M	C	T	S	L
G916a2	WIAF-13211	HT28523	762	Huntingtin associated protein 1-like protein	CGGAGGAGTT[G/C]GTGCCCCAGG	M	G	C	L	F
G916a3	WIAF-13212	HT28523	560	Huntingtin associated protein 1-like protein	GAGCTCAGAA[C/T]GTCTCTAAGG	M	C	T	T	M
G917u1	WIAF-11972	U79734	1075	HIP1, huntingtin interacting protein 1	AGAGCCAGCG[G/A]GTTGTGCTGC	S	G	A	R	R
G917u2	WIAF-11973	U79734	1005	HIP1, huntingtin interacting protein 1	GACCACTTAA[T/C]TGAGCGACTA	M	T	C	I	T
G917u3	WIAF-11977	U79734	1539	HIP1, huntingtin interacting protein 1	CTGCAAGGCA[G/A]CCTGGAAACT	M	G	A	S	N
G917u4	WIAF-12005	U79734	817	HIP1, huntingtin interacting protein 1	TGGTGGTGAT[C/T]CCTGCAGAGG	S	C	T	I	I
G917u5	WIAF-12006	U79734	1906	HIP1, huntingtin interacting protein 1	GCTGGAGCCA[G/C]TATCTGGCCT	M	G	C	Q	H
G917a6	WIAF-13157	U79734	993	HIP1, huntingtin interacting protein 1	AAGGATGAGA[A/G]GGACCACTTA	M	A	G	K	R
G919u1	WIAF-11974	D30742	707	CAMK4, calcium/calmodulin-dependent protein kinase IV	ACTGCGCACC[T/C]GAAATTCTTA	S	T	C	P	P

G919u2	WIAF-11991	D30742	1139	CAMK4, calcium/calmodulin-dependent protein kinase IV	AGAGCCACAA[G/A]GCTAGCCGAG	S	G	A	K	K
G919u3	WIAF-12007	D30742	834	CAMK4, calcium/calmodulin-dependent protein kinase IV	CATGTTCAGG[A/T]GAATTCTGAA	N	A	T	R	*
G919u4	WIAF-13443	D30742	1088	CAMK4, calcium/calmodulin-dependent protein kinase IV	TGGCCTCTTC[C/G]CGCCTGGGAA	S	C	G	S	S
G920u1	WIAF-11979	X78520	1952	CLCN3, chloride channel 3	ATGACATTCC[T/C]GATCGTCCAG	S	T	C	P	P
G920u2	WIAF-11980	X78520	1819	CLCN3, chloride channel 3	ATAGCCTTCC[C/T]TAATCCATAC	M	C	T	P	L
G920u3	WIAF-11981	X78520	2094	CLCN3, chloride channel 3	CATTGGAGCG[A/G]TCGCAGGAAG	M	A	G	I	V
G920u4	WIAF-11983	X78520	2822	CLCN3, chloride channel 3	ATATTTTCCG[A/G]AAGCTGGGAC	S	A	G	R	R
G920u5	WIAF-11984	X78520	2745	CLCN3, chloride channel 3	GCCATTGAAG[C/T]TTCGAAGCAT	M	C	T	L	F
G920u6	WIAF-11987	X78520	2499	CLCN3, chloride channel 3	TCCCTTAGCT[G/T]TCCTGACACA	M	G	T	V	F
G920u7	WIAF-12008	X78520	1251	CLCN3, chloride channel 3	CATCATCAGA[G/A]GTTACTTGGG	M	G	A	G	S
G920u8	WIAF-12011	X78520	888	CLCN3, chloride channel 3	AGTAGTAACA[C/T]TAACAGGATT	S	C	T	L	L
G920u9	WIAF-13459	X78520	2804	CLCN3, chloride channel 3	CAATGGAGAT[T/C]GTGGTGGATA	S	T	C	I	I
G921u1	WIAF-11954	J02908	931	CLU, clusterin (complement lysis inhibitor, SP-40,40, sulfated glycoprotein 2, testosterone-repressed prostate message 2, apolipoprotein J)	GAGAGGTTGA[C/T]CAGGAAATAC	M	C	T	T	I
G921u2	WIAF-11955	J02908	880	CLU, clusterin (complement lysis inhibitor, SP-40,40, sulfated glycoprotein 2, testosterone-repressed prostate message 2, apolipoprotein J)	CCCTCCCAGG[C/T]TAAGCTGCGG	M	C	T	A	V
G921u3	WIAF-11990	J02908	1051	CLU, clusterin (complement lysis inhibitor, SP-40,40, sulfated glycoprotein 2, testosterone-repressed prostate message 2, apolipoprotein J)	CTCACGCAAG[G/C]CGAAGACCAG	M	G	C	G	A

G921u4	WIAF-13469	J02908	986	CLU, clusterin (complement lysis inhibitor, SP-40,40, sulfated glycoprotein 2, testosterone-repressed prostate message 2, apolipoprotein J)	TCAACACCTC[C/T]TCCTTGCTGG	S	C	T	S	S
G923u1	WIAF-11993	M19650	1059	Human 2',3'-cyclic nucleotide 3'-phosphodiesterase mRNA, complete cds.	GAGCTAAGCC[G/A]GGGCAAGCTC	M	G	A	R	Q
G923u2	WIAF-11994	M19650	1062	Human 2',3'-cyclic nucleotide 3'-phosphodiesterase mRNA, complete cds.	CTAAGCCGGG[G/T]CAAGCTCTAT	M	G	T	G	V
G923u3	WIAF-13445	M19650	1141	Human 2',3'-cyclic nucleotide 3'-phosphodiesterase mRNA, complete cds.	TCTTCACGGG[G/A]TACTACGGGA	S	G	A	G	G
G925u1	WIAF-11953	L11315	666	CAK, cell adhesion kinase	GGGTCATGAG[T/C]GTCTGTCTGC	S	T	C	S	S
G925u2	WIAF-11959	L11315	2562	CAK, cell adhesion kinase	TGCTGCCCAT[C/T]CGCTGGATGG	S	C	T	I	I
G925u3	WIAF-11996	L11315	2049	CAK, cell adhesion kinase	AAGATCTGGT[T/C]AGTCTTGATT	S	T	C	V	V
G925u4	WIAF-13440	L11315	1601	CAK, cell adhesion kinase	TACCAGGAGC[C/T]CCGGCCTCGT	M	C	T	P	L
G925u5	WIAF-13441	L11315	1629	CAK, cell adhesion kinase	CGCCCCACTC[C/T]GCTCCCTGTG	S	C	T	S	S
G925u6	WIAF-13451	L11315	2262	CAK, cell adhesion kinase	TGGAGAACGG[C/T]GACCTCAACC	S	C	T	G	G
G926u1	WIAF-11961	AF018956	577	NRP1, neuropilin 1	TGAAAGCTTT[G/T]ACCTGGAGCC	M	G	T	D	Y
G926u2	WIAF-11963	AF018956	1683	NRP1, neuropilin 1	CCACGCGATT[C/G]ATCAGGATCT	M	C	G	F	L
G926u3	WIAF-11975	AF018956	2176	NRP1, neuropilin 1	GACCTTCTGG[T/C]ATCACATGTC	M	T	C	Y	H
G926u4	WIAF-11976	AF018956	2092	NRP1, neuropilin 1	TTCCCAAGCT[G/T]ACGAAAATCA	M	G	T	D	Y
G926a5	WIAF-13158	AF018956	747	NRP1, neuropilin 1	TTTTTTACAC[C/T]GACAGCGCGA	S	C	T	T	T
G926a6	WIAF-13159	AF018956	996	NRP1, neuropilin 1	ACTTGGGCCT[T/C]CTGCGCTTTG	S	T	C	L	L
G926u7	WIAF-13444	AF018956	644	NRP1, neuropilin 1	GAAATCTGGG[A/C]TGGATTCCCT	M	A	C	D	A
G926u8	WIAF-13450	AF018956	1738	NRP1, neuropilin 1	CAGAATGGAG[C/G]TGCTGGGCTG	M	C	G	L	V
G926u9	WIAF-13452	AF018956	537	NRP1, neuropilin 1	TTGTCTTTGC[G/A]CCAAAGATGT	S	G	A	A	A
G926u10	WIAF-13457	AF018956	2197	NRP1, neuropilin 1	TGGGTCCCAC[G/A]TCGGCACACT	M	G	A	V	I
G927u1	WIAF-11978	AF022860	870	NRP2, neuropilin 2	GGATTGCTAA[T/C]GAACAGATCA	S	T	C	N	N
G927u2	WIAF-11982	AF022860	1674	NRP2, neuropilin 2	ATGACACCCC[T/G]GACATCCGAA	S	T	G	P	P
G927u3	WIAF-11985	AF022860	1250	NRP2, neuropilin 2	TGGCACTCAG[G/A]TATCGCCCTC	M	G	A	G	D
G927u4	WIAF-11986	AF022860	1071	NRP2, neuropilin 2	ATGGCTACTA[C/T]GTCAAATCCT	S	C	T	Y	Y

G927u5	WIAF-12009	AF022860	726	NRP2, neuropilin 2	GTTCATCGAC[G/A]GGGATCCTCT	S	G	A	T	T
G927u6	WIAF-12010	AF022860	2522	NRP2, neuropilin 2	GCAACCTCAG[G/T]GTCTGGCGCC	M	G	T	G	V
G927u7	WIAF-12012	AF022860	123	NRP2, neuropilin 2	GCTATATCAC[C/T]TCTCCCGGTT	S	C	T	T	T
G927a8	WIAF-13160	AF022860	2427	NRP2, neuropilin 2	CTTTTGCAGT[G/T]GACATCCCAG	S	G	T	V	V
G927a9	WIAF-13161	AF022860	2430	NRP2, neuropilin 2	TTGCAGTGGA[C/G]ATCCCAGAAA	M	C	G	D	E
G927a10	WIAF-13162	AF022860	2463	NRP2, neuropilin 2	AAGGATATGA[A/G]GATGAAATTG	S	A	G	E	E
G927a11	WIAF-13163	AF022860	2473	NRP2, neuropilin 2	AGATGAAATT[G/T]ATGATGAATA	M	G	T	D	Y
G927u12	WIAF-13480	AF022860	724	NRP2, neuropilin 2	TCGTTCATCG[A/T]CGGGGATCCT	M	A	T	T	S
G927u13	WIAF-13481	AF022860	767	NRP2, neuropilin 2	ATGGCGGTGG[C/T]CAAGGATGGC	M	C	T	A	V
G930a1	WIAF-13164	HT2608	609	GABRA2, gamma-aminobutyric acid (GABA) A receptor, alpha 2	ACAATGGGAA[G/a]AAATCAGTAG	S	G	a	K	K
G931a1	WIAF-13153	HT2609	1111	GABRA3, gamma-aminobutyric acid (GABA) A receptor, alpha 3	ACTGGTTCAT[A/g]GCCGTCTGTT	M	A	g	I	M
G931a2	WIAF-13165	HT2609	1448	GABRA3, gamma-aminobutyric acid (GABA) A receptor, alpha 3	TGTCAGCAAG[G/A]TTGACAAAAT	M	G	A	V	I
G932a1	WIAF-13154	HT27773	1077	GABRA4, gamma-aminobutyric acid (GABA) A receptor, alpha 4	CAAAAGAAAG[A/G]CATCAAAGCC	M	A	G	T	A
G932a2	WIAF-13155	HT27773	1189	GABRA4, gamma-aminobutyric acid (GABA) A receptor, alpha 4	AGAACAAATG[C/A]TTTGGTTCAC	M	C	A	A	D
G936u1	WIAF-12308	HT3432	1027	GABRB2, gamma-aminobutyric acid (GABA) A receptor, beta 2	AATTACGATG[C/T]TTCAGCTGCA	M	C	T	A	V
G936u2	WIAF-12327	HT3432	362	GABRB2, gamma-aminobutyric acid (GABA) A receptor, beta 2	AAGGCTATGA[C/T]ATTCGTCTGA	S	C	T	D	D
G936u3	WIAF-12328	HT3432	571	GABRB2, gamma-aminobutyric acid (GABA) A receptor, beta 2	CTCTGGGTGC[C/T]TGATACCTAT	M	C	T	P	L
G939u1	WIAF-12330	HT2236	1219	GABRR2, gamma-aminobutyric acid (GABA) receptor, rho 2	CTGGATGGAA[G/C]CTACAGTGAG	M	G	C	S	T
G939u2	WIAF-12355	HT2236	1003	GABRR2, gamma-aminobutyric acid (GABA) receptor, rho 2	ACCACCATCA[T/C]CACGGGCGTG	M	T	C	I	T

G939u3	WIAF-12356	HT2236	1041	GABRR2, gamma-aminobutyric acid (GABA) receptor, rho 2	CGTCTCCTAC[G/A]TCAAGGCCGT	M	G	A	V	I									
G950u1	WIAF-13622	U64871	785	Human putative G protein-coupled receptor (GPR19) gene, complete cds.	GTCCTGCTCC[A/C]GTTCACCACT	M	A	C	Q	P									
G950u2	WIAF-13624	U64871	443	Human putative G protein-coupled receptor (GPR19) gene, complete cds.	GATAACAGCA[A/C]GCCACATTTG	M	A	C	K	T									
G950u3	WIAF-13625	U64871	818	Human putative G protein-coupled receptor (GPR19) gene, complete cds.	CTGGGTAGTG[C/T]AACGTGCAAG	M	C	T	A	V									
G955a1	WIAF-13166	HT3860	5110	calcium channel, voltage-gated, alpha 1 subunit, L type, alt. transcript 1	CTGGCCTCTT[T/c]ACCGTGGAGA	S	T	c	F	F									
G955a2	WIAF-13167	HT3860	3842	calcium channel, voltage-gated, alpha 1 subunit, L type, alt. transcript 1	CTACCCCAAC[C/a]CAGAAACTAC	M	C	a	P	T									
G955a3	WIAF-13168	HT3860	5624	calcium channel, voltage-gated, alpha 1 subunit, L type, alt. transcript 1	GTGTGCCCCA[G/a]AGTCCGAGCC	M	G	a	E	K									
G955a4	WIAF-13169	HT3860	5703	calcium channel, voltage-gated, alpha 1 subunit, L type, alt. transcript 1	ATCAGCTTCT[A/g]CATGCTCTGT	M	A	g	Y	C									
G955a5	WIAF-13170	HT3860	5809	calcium channel, voltage-gated, alpha 1 subunit, L type, alt. transcript 1	ACCACCTGGA[T/c]GAGTTTAAAA	S	T	c	D	D									
G955a6	WIAF-13171	HT3860	6616	calcium channel, voltage-gated, alpha 1 subunit, L type, alt. transcript 1	CCGGCTCCAA[C/t]GCCAACATCA	S	C	t	N	N									
G956u1	WIAF-14187	HT2199	1334	calcium channel, voltage-gated, alpha 1D subunit, DHP-sensitive	CTTCACATAG[C/T]CCTTTTGGTA	M	C	T	A	V									
G956u2	WIAF-14188	HT2199	1452	calcium channel, voltage-gated, alpha 1D subunit, DHP-sensitive	AAGAGGACCC[A/T]GCTCCATGTG	S	A	T	P	P									
G956u3	WIAF-14189	HT2199	1614	calcium channel, voltage-gated, alpha 1D subunit, DHP-sensitive	GCTGGACAGA[C/T]GTGCTCTACT	S	C	T	D	D									

G956u4	WIAF-14190	HT2199	2540	calcium channel, voltage-gated, alpha 1D subunit, DHP-sensitive	GGCAAGTTTA[A/T]TTTGATGAA	M	A	T	N	I				
G956u5	WIAF-14191	HT2199	3210	calcium channel, voltage-gated, alpha 1D subunit, DHP-sensitive	TGCTGAGCAG[T/C]GCTGCCCTGG	S	T	C	S	S				
G956u6	WIAF-14192	HT2199	3326	calcium channel, voltage-gated, alpha 1D subunit, DHP-sensitive	TTGAAGATGA[C/T]AACTTTTGGA	M	C	T	T	I				
G956u7	WIAF-14193	HT2199	3274	calcium channel, voltage-gated, alpha 1D subunit, DHP-sensitive	ACTGGGTTAC[T/C]TTGACTATGC	M	T	C	F	L				
G956u8	WIAF-14194	HT2199	5127	calcium channel, voltage-gated, alpha 1D subunit, DHP-sensitive	TGCCTCTCAA[C/T]AGTGACGGGA	S	C	T	N	N				
G956u9	WIAF-14195	HT2199	5173	calcium channel, voltage-gated, alpha 1D subunit, DHP-sensitive	TGCTTTGGTT[C/T]GAACGGCTCT	N	C	T	R	*				
G956u10	WIAF-14200	HT2199	1437	calcium channel, voltage-gated, alpha 1D subunit, DHP-sensitive	CAGATATCGT[A/G]GCTGAAGAGG	S	A	G	V	V				
G956u11	WIAF-14201	HT2199	2567	calcium channel, voltage-gated, alpha 1D subunit, DHP-sensitive	ACCAAGCGGA[G/T]CACCTTTGAC	M	G	T	S	I				
G956u12	WIAF-14202	HT2199	4464	calcium channel, voltage-gated, alpha 1D subunit, DHP-sensitive	TCACCTTTTT[C/T]CGTCTTTTCC	S	C	T	F	F				
G956u13	WIAF-14215	HT2199	6927	calcium channel, voltage-gated, alpha 1D subunit, DHP-sensitive	GCTACAGCGA[C/T]GAAGAGCCAG	S	C	T	D	D				
G956u14	WIAF-14216	HT2199	6858	calcium channel, voltage-gated, alpha 1D subunit, DHP-sensitive	CCCGAGCCAA[C/T]GGGGATGTGG	S	C	T	N	N				
G957u1	WIAF-12306	HT4229	915	calcium channel, voltage-gated, alpha 1E subunit, alt. transcript 2	TACATCGAGC[G/A]TGCTTCATGA	M	G	A	?	R				
G957u2	WIAF-12309	HT4229	3555	calcium channel, voltage-gated, alpha 1E subunit, alt. transcript 2	GCCACTACAT[C/T]GTGAACCTGC	S	C	T	I	I				

G957u3	WIAF-12310	HT4229	4116	calcium channel, voltage-gated, alpha 1E subunit, alt. transcript 2	ATGTAGATCA [C/T] GAGAAAAACA	S	C	T	H	H
G957u4	WIAF-12313	HT4229	5181	calcium channel, voltage-gated, alpha 1E subunit, alt. transcript 2	AGAACGAGAA [T/C] GAACGCTGCG	S	T	C	N	N
G957u5	WIAF-12314	HT4229	5971	calcium channel, voltage-gated, alpha 1E subunit, alt. transcript 2	TATGGACCCC [G/A] CCGATGACGG	S	G	A	T	T
G957u6	WIAF-12315	HT4229	5985	calcium channel, voltage-gated, alpha 1E subunit, alt. transcript 2	ATGACGGACA [G/T] TTCCAAGAAC	M	G	T	Q	H
G957u7	WIAF-12329	HT4229	3100	calcium channel, voltage-gated, alpha 1E subunit, alt. transcript 2	GCTGGCAGGA [G/A] GCCTTGATGA	M	G	A	G	S
G957u8	WIAF-12331	HT4229	6492	calcium channel, voltage-gated, alpha 1E subunit, alt. transcript 2	CCCTCCTTTC [C/T] TACAGCTCCC	M	C	T	?	R
G957u9	WIAF-12354	HT4229	3839	calcium channel, voltage-gated, alpha 1E subunit, alt. transcript 2	AACGCTTTGG [G/C] AACCAACAAA	M	G	C	G	A
G957u10	WIAF-12357	HT4229	4753	calcium channel, voltage-gated, alpha 1E subunit, alt. transcript 2	TGACTTCATC [A/G] CCGTGATTGG	M	A	G	T	A
G960u1	WIAF-12305	HT3336	1246	CACNB3, calcium channel, voltage-dependent, beta 3 subunit	TTGATGCCCT [C/T] TGATGAGGCC	M	C	T	S	F
G960u2	WIAF-12340	HT3336	1288	CACNB3, calcium channel, voltage-dependent, beta 3 subunit	TGGACAGGAT [C/T] TTCACAGCGT	M	C	T	S	F
G960u3	WIAF-12345	HT3336	641	CACNB3, calcium channel, voltage-dependent, beta 3 subunit	AGGCTCTCTT [C/T] GACTTCCTCA	S	C	T	F	F
G960u4	WIAF-12346	HT3336	576	CACNB3, calcium channel, voltage-dependent, beta 3 subunit	CATGCGGCCT [G/A] TGGTGCTGGT	M	G	A	V	M
G961u1	WIAF-12322	U95019	2037	CACNB2, calcium channel, voltage-dependent, beta 2 subunit	ACTCTGCCTA [C/T] GTAGAGCCAA	S	C	T	Y	Y

G961u2	WIAF-12347	U95019	2007	CACNB2, calcium channel, voltage-dependent, beta 2 subunit	CATTTGACTC[G/A]GAAACCCAGG	S	G	A	S	S
G962u1	WIAF-12324	U95020	1423	CACNB4, calcium channel, voltage-dependent, beta 4 subunit	CCAATTGAAA[G/A]ACGAAGTCTA	M	G	A	R	K
G962u2	WIAF-12342	U95020	167	CACNB4, calcium channel, voltage-dependent, beta 4 subunit	GGAGCAGGTT[G/T]AAAAGATCCG	M	G	T	L	F
G962u3	WIAF-12350	U95020	1571	CACNB4, calcium channel, voltage-dependent, beta 4 subunit	ACACTTACAA[A/G]CCCCATAGGA	S	A	G	K	K
G965u1	WIAF-12312	U40583	1276	CHRNA7, cholinergic receptor, nicotinic, alpha polypeptide 7	TCCTGCACGG[T/C]GGGCAACCCC	S	T	C	G	G
G968a1	WIAF-12119	HT27592	1008	CHRNA1, cholinergic receptor, nicotinic, alpha polypeptide 1 (muscle)	ACACACACCA[C/T]CGCTCACCCA	S	C	T	H	H
G968u2	WIAF-12368	HT27592	1136	CHRNA1, cholinergic receptor, nicotinic, alpha polypeptide 1 (muscle)	AAGATTTTTA[C/T]AGAAGACATT	M	C	T	T	I
G973a1	WIAF-13172	HT48774	800	CHRNA2, cholinergic receptor, nicotinic, alpha polypeptide 2 (neuronal)	ACACTTCAGA[C/t]GTGGTGATTG	S	C	t	D	D
G973a2	WIAF-13173	HT48774	927	CHRNA2, cholinergic receptor, nicotinic, alpha polypeptide 2 (neuronal)	CTGGAACCCC[G/a]CTGATTTTGG	M	G	a	A	T
G977u1	WIAF-13949	Y08419	366	CHRNA5, cholinergic receptor, nicotinic, alpha polypeptide 5	AAGTTATACG[T/C]GTTCCTTCAG	S	T	C	R	R
G978a1	WIAF-13179	Y08417	1331	CHRNB3, cholinergic receptor, nicotinic, beta polypeptide 3	CCATTAGATA[C/a]ATTTCGAGAC	N	C	a	Y	*
G983a1	WIAF-13214	HT0374	236	NPY, neuropeptide Y	GATACTACTC[G/A]GCGCTGCGAC	S	G	A	S	S
G983a2	WIAF-13215	HT0374	290	NPY, neuropeptide Y	GAAAACGATC[C/T]AGCCCAGAGA	S	C	T	S	S
G983a3	WIAF-13216	HT0374	111	NPY, neuropeptide Y	GCGACTGGGG[C/T]TGTCCGGACT	S	C	T	L	L
G987a1	WIAF-13174	HT27830	159	PPYR1, pancreatic polypeptide receptor 1	TGGTCTTCAT[C/T]GTCACTTCCT	S	C	T	I	I

G987a2	WIAF-13175	HT27830	222	PPYR1, pancreatic polypeptide receptor 1	TGATGTGTGT[G/A]ACTGTGAGGC	S	G	A	V	V
G987a3	WIAF-13176	HT27830	322	PPYR1, pancreatic polypeptide receptor 1	GCCGCTGACC[G/T]CCGTCTACAC	M	G	T	A	S
G987a4	WIAF-13177	HT27830	1074	PPYR1, pancreatic polypeptide receptor 1	TGGAGGAGTC[G/A]GAGCATCTGC	S	G	A	S	S
G987a5	WIAF-13178	HT27830	975	PPYR1, pancreatic polypeptide receptor 1	CCTCCACCTG[C/T]GTCAACCCAT	S	C	T	C	C
G987a6	WIAF-13180	HT27830	615	PPYR1, pancreatic polypeptide receptor 1	AGTTCCTGGC[A/g]GATAAGGTGG	S	A	g	A	A
G987a7	WIAF-13181	HT27830	718	PPYR1, pancreatic polypeptide receptor 1	GGGCTTCATC[C/T]TGGTCTGTTA	S	C	T	L	L
G987a8	WIAF-13182	HT27830	745	PPYR1, pancreatic polypeptide receptor 1	CATCTACCGG[C/t]GCCTGCAGAG	M	C	t	R	C
G987a9	WIAF-13183	HT27830	842	PPYR1, pancreatic polypeptide receptor 1	GTGATGGTGG[T/A]GGCCTTTGCC	M	T	A	V	E
G987a10	WIAF-13184	HT27830	852	PPYR1, pancreatic polypeptide receptor 1	TGGCCTTTGC[C/T]GTGCTCTGGC	S	C	T	A	A
G987a11	WIAF-13185	HT27830	889	PPYR1, pancreatic polypeptide receptor 1	CAACAGCCTG[G/a]AAGACTGGCA	M	G	a	E	K
G987a12	WIAF-13186	HT27830	924	PPYR1, pancreatic polypeptide receptor 1	CCATCTGCCA[C/T]GGGAACCTCA	S	C	T	H	H
G989u1	WIAF-13573	D86519	891	NPY6R, neuropeptide Y receptor Y6	TGACTCATGC[C/T]TACTGGGGCA	S	C	T	A	A
G989u2	WIAF-13588	D86519	465	NPY6R, neuropeptide Y receptor Y6	ACCACCCAGC[A/G]TCTAATACAA	S	A	G	A	A
G989u3	WIAF-13591	D86519	980	NPY6R, neuropeptide Y receptor Y6	GAGCCCTTCC[G/A]CAACCTCTCT	M	G	A	R	H
G991u1	WIAF-12390	HT97376	336	Notch2	AAGGTACTTG[C/T]GTTCAGAAAA	S	C	T	C	C
G993u1	WIAF-12359	U95299	1343	NOTCH4, Notch (Drosophila) homolog 4	TCCACACTCT[G/T]CCTGTGTCAG	M	G	T	C	F
G993u2	WIAF-12361	U95299	2020	NOTCH4, Notch (Drosophila) homolog 4	TAAGGACCAG[A/G]AAGACAAGGC	M	A	G	K	E
G993u3	WIAF-12384	U95299	5775	NOTCH4, Notch (Drosophila) homolog 4	GGGCCTATTC[G/T]CATTGCCGGA	S	G	T	S	S
G996a1	WIAF-13213	HT3329	356	OPRM1, opioid receptor, mu 1	CTTAGATGGC[A/G]ACCTGTCCGA	M	A	G	N	D
LPLa4	WIAF-13314	HT1320	443	LPL, lipoprotein lipase	ATGTATGAGA[G/T]TTGGGTGCCA	M	G	T	S	I
LPLa5	WIAF-13315	HT1320	579	LPL, lipoprotein lipase	GACAGGATGT[G/A]GCCCGGTTTA	S	G	A	V	V

LPLa6	WIAF-13316	HT1320	609	LPL, lipoprotein lipase	TGGAGGAGGA [G/A] TTTAACTACC	S	G	A	E	E
LPLa7	WIAF-13317	HT1320	1338	LPL, lipoprotein lipase	CAAATAAGAC [C/A] TACTCCTTCC	S	C	A	T	T
LPLa8	WIAF-13318	HT1320	1117	LPL, lipoprotein lipase	CAATCTGGGC [T/G] ATGAGATCAA	M	T	G	Y	D
LPLa9	WIAF-13319	HT1320	715	LPL, lipoprotein lipase	CAGAATTACT [G/A] GCCTCGATCC	M	G	A	G	S
LPLa10	WIAF-13320	HT1320	834	LPL, lipoprotein lipase	CTGGTCGAAG [C/A] ATTGGAATCC	M	C	A	S	R
LPLa11	WIAF-13321	HT1320	951	LPL, lipoprotein lipase	GACTTCGAGA [T/A] GTGGACCAGC	M	T	A	D	E
LPLa12	WIAF-13322	HT1320	1595	LPL, lipoprotein lipase	AATAAGAAGT [C/G] AGGCTGAAAC	N	C	G	S	*
LPLa13	WIAF-13323	HT1320	1597	LPL, lipoprotein lipase	TAAGAAGTCA [G/A] GCTGAAACTG	M	G	A	G	S
LPLa14	WIAF-13324	HT1320	1606	LPL, lipoprotein lipase	AGGCTGAAAC [T/C] GGGCGAATCT	-	T	C	-	-
LPLa15	WIAF-13325	HT1320	1611	LPL, lipoprotein lipase	GAAACTGGGC [G/A] AATCTACAGA	-	G	A	-	-

(This page has been intentionally left blank.)

While this invention has been particularly shown and described with references to preferred embodiments thereof, it will be understood by those skilled in the art that various changes in form and details may be made therein without departing from the scope of the invention encompassed by the appended claims.

CLAIMS

WE CLAIM:

1. A method of diagnosing or aiding in the diagnosis of a vascular disease in an individual comprising
   a) obtaining a nucleic acid sample from the individual; and
   b) determining the nucleotide present at nucleotide position 2210 of the thrombospondin-1 gene,
   wherein presence of a G at nucleotide position 2210 is indicative of increased likelihood of a vascular disease in the individual as compared with an individual having an A at nucleotide position 2210.

2. The method of Claim 1, wherein the thrombospondin-1 gene has the nucleotide sequence of SEQ ID NO: 1.

3. The method of Claim 1, wherein the vascular disease is selected from the group consisting of atherosclerosis, coronary heart disease, myocardial infarction, stroke, peripheral vascular diseases, venous thromboembolism and pulmonary embolism.

4. The method of Claim 3, wherein the vascular disease is myocardial infarction.

5. The method of Claim 3, wherein the vascular disease is coronary heart disease.

6. A method of diagnosing or aiding in the diagnosis of a vascular disease in an individual comprising
   a) obtaining a nucleic acid sample from the individual; and
   b) determining the nucleotide present at nucleotide position 2210 of the thrombospondin-1 gene,

wherein presence of an A at nucleotide position 2210 is indicative of decreased likelihood of a vascular disease in the individual as compared with an individual having a G at nucleotide position 2210.

7. The method according to Claim 6, wherein the thrombospondin-1 gene has the nucleotide sequence of SEQ ID NO: 1.

8. The method according to Claim 6, wherein the vascular disease is selected from the group consisting of atherosclerosis, coronary heart disease, myocardial infarction, stroke, peripheral vascular diseases, venous thromboembolism and pulmonary embolism.

9. The method according to Claim 8, wherein the vascular disease is myocardial infarction.

10. The method according to Claim 8, wherein the vascular disease is coronary heart disease.

11. A method for predicting the likelihood that an individual will have a vascular disease, comprising the steps of:

a) obtaining a DNA sample from an individual to be assessed; and

b) determining the nucleotide present at nucleotide position 2210 of the thrombospondin-1 gene,

wherein presence of a G at nucleotide position 2210 is indicative of increased likelihood of a vascular disease in the individual as compared with an individual having an A at nucleotide position 2210.

12. The method according to Claim 11, wherein the thrombospondin-1 gene has the nucleotide sequence of SEQ ID NO: 1.

13. The method according to Claim 11, wherein the individual is an individual at risk for development of a vascular disease.

14. The method according to Claim 11, wherein the vascular disease is selected from the group consisting of atherosclerosis, coronary heart disease, myocardial infarction, stroke, peripheral vascular diseases, venous thromboembolism and pulmonary embolism.

15. The method according to Claim 14, wherein the vascular disease is myocardial infarction.

16. The method according to Claim 14, wherein the vascular disease is coronary heart disease.

17. A nucleic acid molecule comprising all or a portion of the nucleic acid sequence of SEQ ID NO: 1 wherein said nucleic acid molecule is at least 10 nucleotides in length and wherein the nucleic acid sequence comprises a polymorphic site at nucleotide position 2210 of SEQ ID NO: 1.

18. The nucleic acid molecule according to Claim 17, wherein the nucleotide at the polymorphic site is different from a nucleotide at the polymorphic site in a corresponding reference allele.

19. An allele-specific oligonucleotide that hybridizes to the nucleic acid molecule of Claim 17.

20. A peptide of SEQ ID NO: 2 which is at least ten contiguous amino acids, wherein the peptide comprises the serine at amino acid position 700 of SEQ ID NO: 2.

21. A method of diagnosing or aiding in the diagnosis of a vascular disease in an individual comprising

a) obtaining a biological sample comprising thrombospondin-1 protein or relevant portion thereof from the individual; and

b) determining the amino acid present at amino acid position 700 of the thrombospondin-1 protein,

wherein presence of an asparagine at amino acid position 700 is indicative of increased likelihood of a vascular disease in the individual as compared with an individual having a serine at amino acid position 700.

22. The method of Claim 21, wherein the thrombospondin-1 protein has the amino acid sequence of SEQ ID NO: 2.

23. The method of Claim 22, wherein the vascular disease is selected from the group consisting of atherosclerosis, coronary heart disease, myocardial infarction, stroke, peripheral vascular diseases, venous thromboembolism and pulmonary embolism.

24. The method of Claim 23, wherein the vascular disease is myocardial infarction.

25. The method of Claim 23, wherein the vascular disease is coronary heart disease.

26. A method of diagnosing or aiding in the diagnosis of a vascular disease in an individual comprising

a) obtaining a biological sample comprising thrombospondin-1 protein or relevant portion thereof from the individual; and

b) determining the amino acid present at amino acid position 700 of the thrombospondin-1 protein,

wherein presence of a serine at amino acid position 700 is indicative of reduced likelihood of a vascular disease in the individual as compared with an individual having an asparagine at amino acid position 700.

27. The method according to Claim 26, wherein the thrombospondin-1 protein has the amino acid sequence of SEQ ID NO: 2.

28. The method according to Claim 26, wherein the vascular disease is selected from the group consisting of atherosclerosis, coronary heart disease, myocardial infarction, stroke, peripheral vascular diseases, venous thromboembolism and pulmonary embolism.

29. The method of Claim 28, wherein the vascular disease is myocardial infarction.

30. The method of Claim 28, wherein the vascular disease is coronary heart disease.

31. A method of diagnosing or aiding in the diagnosis of a vascular disease in an individual comprising

a) obtaining a nucleic acid sample from the individual; and

b) determining the nucleotide present at nucleotide position 1186 of the thrombospondin-4 gene,

wherein presence of a C at nucleotide position 1186 is indicative of increased likelihood of a vascular disease in the individual as compared with an individual having an G at nucleotide position 1186.

32. The method of Claim 31, wherein the thrombospondin-4 gene has the nucleotide sequence of SEQ ID NO: 3.

33. The method of Claim 31, wherein the vascular disease is selected from the group consisting of atherosclerosis, coronary heart disease, myocardial infarction, stroke, peripheral vascular diseases, venous thromboembolism and pulmonary embolism.

34. The method of Claim 33, wherein the vascular disease is myocardial infarction.

35. The method of Claim 33, wherein the vascular disease is coronary heart disease.

36. A method of diagnosing or aiding in the diagnosis of a vascular disease in an individual comprising

a) obtaining a nucleic acid sample from the individual; and

b) determining the nucleotide present at nucleotide position 1186 of the thrombospondin-4 gene,

wherein presence of a G at nucleotide position 1186 is indicative of decreased likelihood of a vascular disease in the individual as compared with an individual having a C at nucleotide position 1186.

37. The method according to Claim 36, wherein the thrombospondin-4 gene has the nucleotide sequence of SEQ ID NO: 3.

38. The method according to Claim 36, wherein the vascular disease is selected from the group consisting of atherosclerosis, coronary heart disease, myocardial infarction, stroke, peripheral vascular diseases, venous thromboembolism and pulmonary embolism.

39. The method according to Claim 38, wherein the vascular disease is myocardial infarction.

40. The method according to Claim 38, wherein the vascular disease is coronary heart disease.

41. A method for predicting the likelihood that an individual will have a vascular disease, comprising the steps of:

a) obtaining a DNA sample from an individual to be assessed; and

b) determining the nucleotide present at nucleotide position 1186 of the thrombospondin-4 gene,

wherein presence of a C at nucleotide position 1186 is indicative of increased likelihood of a vascular disease in the individual as compared with an individual having a G at nucleotide position 1186.

42. The method according to Claim 41, wherein the thrombospondin-4 gene has the nucleotide sequence of SEQ ID NO: 3.

43. The method according to Claim 41, wherein the individual is an individual at risk for development of a vascular disease.

44. The method according to Claim 41, wherein the vascular disease is selected from the group consisting of atherosclerosis, coronary heart disease, myocardial infarction, stroke, peripheral vascular diseases, venous thromboembolism and pulmonary embolism.

45. The method according to Claim 44, wherein the vascular disease is myocardial infarction.

46. The method according to Claim 44, wherein the vascular disease is coronary heart disease.

47. A nucleic acid molecule comprising all or a portion of the nucleic acid sequence of SEQ ID NO: 3 wherein said nucleic acid molecule is at least 10 nucleotides in length and wherein the nucleic acid sequence comprises a polymorphic site at nucleotide position 1186 of SEQ ID NO: 3.

48. The nucleic acid molecule according to Claim 47, wherein the nucleotide at the polymorphic site is different from a nucleotide at the polymorphic site in a corresponding reference allele.

49. An allele-specific oligonucleotide that hybridizes to the nucleic acid molecule of Claim 47.

50. A peptide of SEQ ID NO: 4 which is at least ten contiguous amino acids, wherein the peptide comprises the proline at amino acid position 387 of SEQ ID NO: 4.

51. A method of diagnosing or aiding in the diagnosis of a vascular disease in an individual comprising

a) obtaining a biological sample comprising thrombospondin-4 protein or relevant portion thereof from the individual; and

b) determining the amino acid present at amino acid position 387 of the thrombospondin-4 protein,

wherein presence of an alanine at amino acid position 387 is indicative of increased likelihood of a vascular disease in the individual as compared with an individual having a proline at amino acid position 387.

52. The method of Claim 51, wherein the thrombospondin-4 protein has the amino acid sequence of SEQ ID NO: 4.

53. The method of Claim 52, wherein the vascular disease is selected from the group consisting of atherosclerosis, coronary heart disease, myocardial infarction, stroke, peripheral vascular diseases, venous thromboembolism and pulmonary embolism.

54. The method of Claim 53, wherein the vascular disease is myocardial infarction.

55. The method of Claim 53, wherein the vascular disease is coronary heart disease.

56. A method of diagnosing or aiding in the diagnosis of a vascular disease in an individual comprising

a) obtaining a biological sample comprising thrombospondin-4 protein or relevant portion thereof from the individual; and

b) determining the amino acid present at amino acid position 387 of the thrombospondin-4 protein,

wherein presence of a proline at amino acid position 387 is indicative of reduced likelihood of a vascular disease in the individual as compared with an individual having an alanine at amino acid position 387.

57. The method according to Claim 56, wherein the thrombospondin-4 protein has the amino acid sequence of SEQ ID NO: 4.

58. The method according to Claim 56, wherein the vascular disease is selected from the group consisting of atherosclerosis, coronary heart disease, myocardial infarction, stroke, peripheral vascular diseases, venous thromboembolism and pulmonary embolism.

59. The method of Claim 58, wherein the vascular disease is myocardial infarction.

60. The method of Claim 58, wherein the vascular disease is coronary heart disease.

61. A nucleic acid molecule selected from the group consisting of the genes listed in the Table, wherein said nucleic acid molecule is at least 10 nucleotides in length and comprises a polymorphic site identified in the Table, wherein a nucleotide at the polymorphic site is different from a nucleotide at the polymorphic site in a corresponding reference allele.

62. A nucleic acid molecule according to Claim 61, wherein said nucleic acid molecule is at least 15 nucleotides in length.

63. A nucleic acid molecule according to Claim 61, wherein said nucleic acid molecule is at least 20 nucleotides in length.

64. A nucleic acid molecule according to Claim 61, wherein the nucleotide at the polymorphic site is the variant nucleotide for the gene listed in the Table.

65. An allele-specific oligonucleotide that hybridizes to a portion of a gene selected from the group consisting of the genes listed in the Table, wherein said portion is at least 10 nucleotides in length and comprises a polymorphic site identified in the Table, wherein a nucleotide at the polymorphic site is different from a nucleotide at the polymorphic site in a corresponding reference allele.

66. An allele-specific oligonucleotide according to Claim 65 that is a probe.

67. An allele-specific oligonucleotide according to Claim 65, wherein a central position of the probe aligns with the polymorphic site of the portion.

68. An allele-specific oligonucleotide according to Claim 65 that is a primer.

69. An allele-specific oligonucleotide according to Claim 68, wherein the 3' end of the primer aligns with the polymorphic site of the portion.

70. An isolated gene product encoded by a nucleic acid molecule according to Claim 61.

71. A method of analyzing a nucleic acid sample, comprising obtaining the nucleic acid sample from an individual; and determining a base occupying any one of the polymorphic sites shown in the Table.

72. A method according to Claim 71, wherein the nucleic acid sample is obtained from a plurality of individuals, and a base occupying one of the polymorphic

positions is determined in each of the individuals, and wherein the method further comprising testing each individual for the presence of a disease phenotype, and correlating the presence of the disease phenotype with the base.

HT1220 Report

RECORD INFORMATION

Gene ID:	1220
Sequence ID:	1220
Protein ID:	1220
Sequence name:	thrombospondin 1, alt. transcript 1
Genome:	nucleus
Taxon:	Homo sapiens
Locus:	1220
Common Name:	thrombospondin 1
Role ID:	40
Coding sequence length:	3513 nt
Transcript sequence length:	5722 nt
Expression data:	481987

ACCESSION DATA

HT1220 is derived from accessions(s):

SP:P07996 (THROMBOSPONDIN 1 PRECURSOR.)
GB:X04665(Human mRNA for thrombospondin)
GB:X14787(Human mRNA for thrombospondin)
GB:U12471(thrombospondin-p50 {Homo sapiens})
GB:M99425(Human thrombospondin mRNA, 3' end.)
PIR:G01478 (thrombospondin-p50 - human (fragment))
GB:U12471(Human thrombospondin-1 gene, partial cds.)
GB:J04835(Human thrombospondin gene, exons 1, 2 and 3.)
GB:M25631(Homo sapiens (clone lambda-TS-33) thrombospondin (THBS) mRNA, 5' end.)

ALTERNATIVE SPLICE INFORMATION

Alternative splice forms for this gene:

HT3987 thrombospondin 1, alt. transcript 2

MAPPING DATA

GDB accession(s) for this gene:

GDB ID:	Symbol

Figure 1A

gdb:120438 THBS1

cDNA FEATURES

Feature	End 5	End 3
coding_seq	112	3624
3'UT	3625	5722
spjunc_h	1235	1236

SEQUENCE

nucleotide:

ggacgcacaggcattccccgcgcccctccagccctcgccgccctcgccaccgctcccggc
cgccgcgctccggtacacacaggatccctgctgggcaccaacagctccaccatggggctg
gcctggggactaggcgtcctgttcctgatgcatgtgtggcaccaaccgcattccagag
tctggcggagacaacagcgtgtttgacatctttgaactcaccggggccgcccgcaagggg
tctgggcgccgactggtgaagggccccgacccttccagcccagctttccgcatcgaggat
gccaacctgatcccccctgtgcctgatgacaagttccaagacctggtggatgctgtgcgg
gcagaaaagggtttcctccttctggcatccctgaggcagatgaagaagacccggggcacg
ctgctggccctggagcggaaagaccactctggccaggtcttcagcgtggtgtccaatggc
aaggcgggcaccctggacctcagcctgaccgtccaaggaaagcagcacgtggtgtctgtg
gaagaagctctcctggcaaccggccagtggaagagcatcaccctgtttgtgcaggaagac
agggcccagctgtacatcgactgtgaaaagatggagaatgctgagttggacgtccccatc
caaagcgtcttcaccagagacctggccagcatcgccagactccgcatcgcaaaggggggc
gtcaatgacaatttccaggggtgctgcagaatgtgaggtttgtctttggaaccacacca
gaagacatcctcaggaacaaaggctgctccagctctaccagtgtcctcctcacccttgac
aacaacgtggtgaatggttccagccctgccatccgcactaactacattggccacaagaca
aaggacttgcaagccatctgcggcatctcctgtgatgagctgtccagcatggtcctggaa
ctcaggggcctgcgcaccattgtgaccacgctgcaggacagcatccgcaaagtgactgaa
gagaacaaagagttggccaatgagctgaggcggcctcccctatgctatcacaacggagtt
cagtacagaaataacgaggaatggactgttgatagctgcactgagtgtcactgtcagaac
tcagttaccatctgcaaaaaggtgtcctgccccatcatgccctgctccaatgccacagtt
cctgatggagaatgctgtcctcgctgttggcccagcgactctgcggacgatggctggtct
ccatggtccgagtggacctcctgttctacgagctgtggcaatggaattcagcagcgcggc
cgctcctgcgatagcctcaacaaccgatgtgagggctcctcggtccagacacggacctgc
cacattcaggagtgtgacaaaagatttaaacaggatggtggctggagccactggtccccg
tggtcatcttgttctgtgacatgtggtgatggtgtgatcacaaggatccggctctgcaac
tctcccagcccccagatgaatgggaaaccctgtgaaggcgaagcgcgggagaccaaagcc
tgcaagaaagacgcctgccccatcaatggaggctggggtccttggtcaccatgggacatc
tgttctgtcacctgtggaggaggggtacagaaacgtagtcgtctctgcaacaaccccgca
ccccagtttggaggcaaggactgcgttggtgatgtaacagaaaaccagatctgcaacaag
caggactgtccaattgatggatgcctgtccaatccctgctttgccggcgtgaagtgtact
agctaccctgatggcagctggaaatgtggtgcttgtccccctggttacagtggaaatggc
atccagtgcacagatgttgatgagtgcaaagaagtgcctgatgcctgcttcaaccacaat
ggagagcaccggtgtgagaacacggaccccggctacaactgcctgccctgcccccacgc
ttcaccggctcacagcccttcggccagggtgtcgaacatgccacggccaacaaacaggtg
tgcaagccccgtaacccctgcacggatgggacccacgactgcaacaagaacgccaagtgc
aactacctgggccactatagcgaccccatgtaccgctgcgagtgcaagcctggctacgct
ggcaatggcatcatctgcggggaggacacagacctggatggctggcccaatgagaacctg
gtgtgcgtggccaatgcgacttaccactgcaaaaaggataattgccccaaccttcccaac
tcagggcaggaagactatgacaaggatggaattggtgatgcctgtgatgatgacgatgac
aatgataaaattccagatgacagggacaactgtccattccattacaacccagctcagtat
gactatgacagagatgatgtgggagaccgctgtgacaactgtccctacaaccacaaccca

Figure 1B

gatcaggcagacacagacaacaatggggaaggagacgcctgtgctgcagacattgatgga
gacggtatcctcaatgaacgggacaactgccagtacgtctacaatgtggaccagagagac
actgatatggatggggttggagatcagtgtgacaattgccccttggaacacaatccggat
cagctggactctgactcagaccgcattggagatacctgtgacaacaatcaggatattgat
gaagatggccaccagaacaatctggacaactgtccctatgtgcccaatgccaaccaggct
gaccatgacaaagatggcaagggagatgcctgtgaccacgatgatgacaacgatggcatt
cctgatgacaaggacaactgcagactcgtgcccaatcccgaccagaaggactctgacggc
gatggtcgaggtgatgcctgcaaagatgattttgaccatgacagtgtgccagacatcgat
gacatctgtcctgagaatgttgacatcagtgagaccgatttccgccgattccagatgatt
cctctggaccccaaagggacatcccaaaatgaccctaactgggttgtacgccatcagggt
aaagaactcgtccagactgtcaactgtgatcctggactcgctgtaggttatgatgagttt
aatgctgtggacttcagtggcaccttcttcatcaacaccgaaagggacgatgactatgct
ggatttgtctttggctaccagtccagcagccgcttttatgttgtgatgtggaagcaagtc
acccagtcctactgggacaccaaccccacgagggctcagggatactcgggcctttctgtg
aaagttgtaaactccaccacagggcctggcgagcacctgcggaacgccctgtggcacaca
ggaaacacccctggccaggtgcgcaccctgtggcatgaccctcgtcacataggctggaaa
gatttcaccgcctacagatggcgtctcagccacaggccaaagacgggtttcattagagtg
gtgatgtatgaagggaagaaaatcatggctgactcaggacccatctatgataaaacctat
gctggtggtagactagggttgtttgtcttctctcaagaaatggtgttcttctctgacctg
aaatacgaatgtagagatccctaatcatcaaattgttgattgaaagactgatcataaacc
aatgctggtattgcaccttctggaactatgggcttgagaaaacccccaggatcacttctc
cttggcttccttctttttctgtgcttgcatcagtgtggactcctagaacgtgcgacctgcc
tcaagaaaatgcagttttcaaaaacagactcatcagcattcagcctccaatgaataagac
atcttccaagcatataaacaattgctttggtttccttttgaaaaagcatctacttgcttc
agttgggaaggtgcccattccactctgcctttgtcacagagcagggtgctattgtgaggc
catctctgagcagtggactcaaaagcattttcaggcatgtcagagaagggaggactcact
agaattagcaaacaaaaccaccctgacatcctccttcaggaacacggggagcagaggcca
aagcactaaggggagggcgcatacccgagacgattgtatgaagaaaatatggaggaactg
ttacatgttcggtactaagtcatttttcaggggattgaaagactattgctggatttcatga
tgctgactggcgttagctgattaacccatgtaaataggcacttaaatagaagcaggaaag
ggagacaaagactggcttctggacttcctccctgatccccacccttactcatcaccttgc
agtggccagaattagggaatcagaatcaaaccagtgtaaggcagtgctggctgccattgc
ctggtcacattgaaattggtggcttcattctagatgtagcttgtgcagatgtagcaggaa
aataggaaaacctaccatctcagtgagcaccagctgcctcccaaaggaggggcagccgtg
cttatatttttatggttacaatggcacaaaattattatcaacctaactaaaacattcctt
ttctctttttttccgtaattactaggtagttttctaattctctcttttggaagtatgattt
ttttaaagtctttacgatgtaaaatatttatttttttacttattctggaagatctggctga
aggattattcatggaacaggaagaagcgtaaagactatccatgtcatctttgttgagagt
cttcgtgactgtaagattgtaaatacagattatttattaactctgttctgcctggaaatt
taggcttcatacggaaagtgtttgagagcaagtagttgacatttatcagcaaatctcttg
caagaacagcacaaggaaaatcagtctaataagctgctctgccccttgtgctcagagtgg
atgttatgggattcctttttttctctgttttatcttttcaagtggaattagttggttatcc
atttgcaaatgttttaaattgcaaagaaagccatgaggtcttcaatactgttttacccca
tcccttgtgcatatttccagggagaaggaaagcatatacacttttttctttcatttttcc
aaaagagaaaaaaatgacaaaaggtgaaacttacatacaaatattacctcatttgttgtg
tgactgagtaaagaatttttggatcaagcggaaagagtttaagtgtctaacaaacttaaa
gctactgtagtacctaaaaagtcagtgttgtacatagcataaaaactctgcagagaagta
ttcccaataaggaaatagcattgaaatgttaaatacaatttctgaaagttatgttttttt
tctatcatctggtataccattgctttatttttataaattattttctcattgccattggaa
tagaatattcagattgtgtagatatgctatttaaataatttatcaggaaatactgcctgt
agagttagtatttctatttttatataatgtttgcacactgaattgaagaattgttggttt
tttctttttttgttttttttttttttttttttttttgctttgacctcccatttttta
ctatttgccaatacctttttctaggaatgtgctttttttgtacacatttttatccattt
tacattctaaagcagtgtaagttgtatattactgtttcttatgtacaaggaacaacaata
aatcatatggaaatttatattt

protein:

MGLAWGLGVLFLMHVCGTNRIPESGGDNSVFDIFELTGAARKGSGRRLVKGPDPSSPAFR

Figure 1C

IEDANLIPPVPDDKFQDLVDAVRAEKGFLLLASLRQMKKTRGTLLALERKDHSGQVFSVV
SNGKAGTLDLSLTVQGKQHVVSVEEALLATGQWKSITLFVQEDRAQLYIDCEKMENAELD
VPIQSVFTRDLASIARLRIAKGGVNDNFQGVLQNVRFVFGTTPEDILRNKGCSSSTSVLL
TLDNNVVNGSSPAIRTNYIGHKTKDLQAICGISCDELSSMVLELRGLRTIVTTLQDSIRK
VTEENKELANELRRPPLCYHNGVQYRNNEEWTVDSCTECHCQNSVTICKKVSCPIMPCSN
ATVPDGECCPRCWPSDSADDGWSPWSEWTSCSTSCGNGIQQRGRSCDSLNNRCEGSSVQT
RTCHIQECDKRFKQDGGWSHWSPWSSCSVTCGDGVITRIRLCNSPSPQMNGKPCEGEARE
TKACKKDACPINGGWGPWSPWDICSVTCGGGVQKRSRLCNNPAPQFGGKDCVGDVTENQI
CNKQDCPIDGCLSNPCFAGVKCTSYPDGSWKCGACPPGYSGNGIQCTDVDECKEVPDACF
NHNGEHRCENTDPGYNCLPCPPRFTGSQPFGQGVEHATANKQVCKPRNPCTDGTHDCNKN
AKCNYLGHYSDPMYRCECKPGYAGNGIICGEDTDLDGWPNENLVCVANATYHCKKDNCPN
LPNSGQEDYDKDGIGDACDDDDDNDKIPDDRDNCPFHYNPAQYDYDRDDVGDRCDNCPYN
HNPDQADTDNNGEGDACAADIDGDGILNERDNCQYVYNVDQRDTDMDGVGDQCDNCPLEH
NPDQLDSDSDRIGDTCDNNQDIDEDGHQNNLDNCPYVPNANQADHDKDGKGDACDHDDDN
DGIPDDKDNCRLVPNPDQKDSDGDGRGDACKDDFDHDSVPDIDDICPENVDISETDFRRF
QMIPLDPKGTSQNDPNWVVRHQGKELVQTVNCDPGLAVGYDEFNAVDFSGTFFINTERDD
DYAGFVFGYQSSSRFYVVMWKQVTQSYWDTNPTRAQGYSGLSVKVVNSTTGPGEHLRNAL
WHTGNTPGQVRTLWHDPRHIGWKDFTAYRWRLSHRPKTGFIRVVMYEGKKIMADSGPIYD
KTYAGGRLGLFVFSQEMVFFSDLKYECRDP

Figure 1D

HT2143 Report

RECORD INFORMATION

Gene ID:	2081
Sequence ID:	2143
Protein ID:	2125
Sequence name:	thrombospondin 4
Genome:	nucleus
Taxon:	Homo sapiens
Locus:	2081
Common Name:	thrombospondin 4
Role ID:	40
Coding sequence length:	2886 nt
Transcript sequence length:	3074 nt
Expression data:	THC168897

ACCESSION DATA

HT2143 is derived from accessions(s):

SP:P35443 (THROMBOSPONDIN 4 PRECURSOR.)
GB:Z19585(thrombospondin-4 (Homo sapiens))
GB:Z19585(H.sapiens mRNA for thrombospondin-4)
PIR:A55710 (thrombospondin 4 precursor - human)

cDNA FEATURES

Feature	End 5	End 3
coding_seq	28	2913
3'UT	2914	3074

SEQUENCE

nucleotide:

gaattccggggagcaggaagagccaacatgctggccccgcgcggagccgccgtcctcctg
ctgcacctggtcctgcagcggtggctagcggcaggcgcccaggccacccccccaggtcttt
gacttctcccatcttccagtcagaggctaaacccaggcgctctgctgccagtcctgaca
gaccccgccctgaatgatctctatgtgatttccaccttcaagctgcagactaaaagttca
gccaccatcttcggtctttactcttcaactgacaacagtaaatattttgaatttactgtg
atgggacgcttaagcaaagccatcctccgttacctgaagaacgatgggaaggtgcatttg

Figure 2A

gaattccggggagcaggaagagccaacatgctggccccgcgcggagccgccgtcctcctg
ctgcacctggtcctgcagcggtggctagcggcaggcgcccaggccacccccaggtcttt
gaccttctcccatcttccagtcagaggctaaacccaggcgctctgctgccagtcctgaca
gaccccgccctgaatgatctctatgtgatttccaccttcaagctgcagactaaaagttca
gccaccatcttcggtctttactcttcaactgacaacagtaaatattttgaatttactgtg
atgggacgcttaagcaaagccatcctccgttacctgaagaacgatgggaaggtgcatttg
gtggttttcaacaacctgcagctggcagacggaaggcggcacaggatcctcctgaggctg
agcaatttgcagcgaggggccggctccctagagctctacctggactgcatccaggtggat
tccgttcacaatctccccagggcctttgctggcccctcccagaaacctgagaccattgaa
ttgaggactttccagaggaagccacaggacttcttggaagagctgaagctggtggtgaga
ggctcactgttccaggtggccagcctgcaagactgcttcctgcagcagagtgagccactg
gctgccacaggcacaggggactttaaccggcagttcttgggtcaaatgacacaattaaac
caactcctgggagaggtgaaggaccttctgagacagcaggttaaggaaacatcatttttg
cgaaacaccatagctgaatgccaggcttgcggtcctctcaagtttcagtctccgacccca
agcacggtggtcgccccggctccccctgcaccgccaacacgcccacctcgtcggtgtgac
tccaacccatgtttccgaggtgtccaatgtaccgacagtagagatggcttccagtgtggg
ccctgccccgagggctacacaggaaacgggatcacctgtattgatgttgatgagtgcaaa
taccatccctgctacccgggcgtgcactgcataaatttgtctcctggcttcagatgtgac
gcctgcccagtgggcttcacagggcccatggtgcagggtgttgggatcagttttgccaag
tcaaacaagcaggtctgcactgacattgatgagtgtcgaaatggagcgtgcgttcccaac
tcgatctgcgttaatactttgggatcttaccgctgtgggccttgtaagccggggtatact
ggtgatcagataaggggatgcaaagtggaaagaaactgcagaaacccagagctgaaccct
tgcagtgtgaatgcccagtgcattgaagagaggcaggggatgtgacatgtgtgtgtgga
gtcggttgggctggagatggctatatctgtggaaaggatgtggacatcgacagttacccc
gacgaagaactgccatgctctgccaggaactgtaaaaaggacaactgcaaatatgtgcca
aattctggccaagaagatgcagacagagatggcattggcgacgcttgtgacgaggatgct
gacggagatgggatcctgaatgagcaggataactgtgtcctgattcataatgtggaccaa
aggaacagcgataaagatatctttggggatgcctgtgataactgcctgagtgtcttaaat
aacgaccagaaagacaccgatggggatggaagaggagatgcctgtgatgatgacatggat
ggagatggaataaaaaacattctggacaactgcccaaaatttcccaatcgtgaccaacgg
gacaaggatggtgatggtgtgggggatgcctgtgacagttgtcctgatgtcagcaaccct
aaccagtctgatgtggataatgatctggttggggactcctgtgacaccaatcaggacagt
gatggagatgggcaccaggacagcacagacaactgccccaccgtcattaacagtgcccag
ctggacaccgataaggatggaattggtgacgagtgtgatgatgatgatgacaatgatggt
atcccagacctggtgccccctggaccagacaactgccggctggtccccaacccagcccag
gaggatagcaacagcgacggagtgggagacatctgtgagtctgactttgaccaggaccag
gtcatcgatcggatcgacgtctgcccagagaacgcagaggtcaccctgaccgacttcagg
gcttaccagaccgtgggcctggatcctgaaggggatgcccagatcgatcccaactgggtg
gtcctgaaccagggcatggagattgtacagaccatgaacagtgatcctggcctggcagtg
gggtacacagcttttaatggagttgacttcgaagggaccttccatgtgaatacccagaca
gatgatgactatgcaggctttatctttggctaccaagatagctccagcttctacgtggtc
atgtggaagcagacggagcagacatattggcaagccaccccattccgagcagttgcagaa
cctggcattcagctcaaggctgtgaagtctaagacaggtccaggggagcatctccggaac
tccctgtggcacacgggggacaccagtgaccaggtcaggctgctgtggaaggactccagg
aatgtgggctggaaggacaaggtgtcctaccgctggttcctacagcacaggccccaggtg
ggctacatcagggtacgattttatgaaggctctgagttggtggctgactctggcgtcacc
atagacaccacaatgcgtggaggccgacttggcgttttctgcttctctcaagaaaacatc
atctggtccaacctcaagtatcgctgcaatgacaccatccctgaggacttccaagagttt
caaacccagaatttcgaccgcttcgataattaaaccaaggaagcaatctgtaactgcttt
tcggaacactaaaaccatatatattttaacttcaattttctttagcttttaccaacccaa
atatatcaaaacgttttatgtgaatgtggcaataaaggagaagagatcatttttaaaaaa
aaaaaaaaaaaaaa

protein:

MLAPRGAAVLLLHLVLQRWLAAGAQATPQVFDLLPSSSQRLNPGALLPVLTDPALNDLYV
ISTFKLQTKSSATIFGLYSSTDNSKYFEFTVMGRLSKAILRYLKNDGKVHLVVFNNLQLA
DGRRHRILLRLSNLQRGAGSLELYLDCIQVDSVHNLPRAFAGPSQKPETIELRTFQRKPQ

Figure 2B

ACDSCPDVSNPNQSDVDNDLVGDSCDTNQDSDGDGHQDSTDNCPTVINSAQLDTDKDGIG
DECDDDDDNDGIPDLVPPGPDNCRLVPNPAQEDSNSDGVGDICESDFDQDQVIDRIDVCP
ENAEVTLTDFRAYQTVGLDPEGDAQIDPNWVVLNQGMEIVQTMNSDPGLAVGYTAFNGVD
FEGTFHVNTQTDDDYAGFIFGYQDSSSFYVVMWKQTEQTYWQATPFRAVAEPGIQLKAVK
SKTGPGEHLRNSLWHTGDTSDQVRLLWKDSRNVGWKDKVSYRWFLQHRPQVGYIRVRFYE
GSELVADSGVTIDTTMRGGRLGVFCFSQENIIWSNLKYRCNDTIPEDFQEFQTQNFDRFD
N

Figure 2C

Poly ID	Sequence ID	Position	Gene Description	Flanking Seq	Mutation Type	Ref NT	Alt NT	Ref AA	Alt AA
G334u4	HT:HT1220_mRNA	2110	THBS1, thrombospondin 1	TGGATGGCTGGCCCA[A/G]TGA GAACCTGGTGTG	Missense	A	G	N	S
G355u2	HT:HT2143_mRNA	1186	THBS4, thrombospondin 4	GAGTGTCGAAATGGA[G/C]CGT GCGTTCCCAACT	Missence	G	C	A	P

Figure 3

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
15 March 2001 (15.03.2001)

PCT

(10) International Publication Number
WO 01/018250 A3

(51) International Patent Classification[7]: C12Q 1/68, C07K 14/47, 14/78

(21) International Application Number: PCT/US00/24503

(22) International Filing Date:
7 September 2000 (07.09.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/153,357	10 September 1999 (10.09.1999)	US
60/220,947	26 July 2000 (26.07.2000)	US
60/225,724	16 August 2000 (16.08.2000)	US

(71) Applicants: **WHITEHEAD INSTITUTE FOR BIOMEDICAL RESEARCH** [US/US]; Nine Cambridge Center, Cambridge, MA 02142 (US). **MILLENNIUM PHARMACEUTICALS, INC.** [US/US]; 75 Sidney Street, Cambridge, MA 02139 (US).

(72) Inventors: **LANDER, Eric, S.**; 151 Bishop Allen Drive, Cambridge, MA 02139 (US). **GARGILL, Michele**; 1 Warwick Park #3, Cambridge, MA 02140 (US). **IRELAND, James, S.**; 1 Warwick Park #3, Cambridge, MA 02140 (US). **BOLK, Stacey**; 202 Baker Street #1, West Roxbury, MA 02132 (US). **DALEY, George, Q.**; 50 Young Road, Weston, MA 02193 (US). **MCCARTHY, Jeanette, J.**; 3625 Dupont Street, San Diego, CA 92106 (US).

(74) Agent: **TREANNIE, Lisa, M.**; Hamilton, Brook, Smith & Reynolds, P.C., 530 Virginia Road, P.O. Box 9133, Concord, MA 01742-9133 (US).

(81) Designated States *(national)*: AU, CA, JP, MX.

(84) Designated States *(regional)*: European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

(88) Date of publication of the international search report:
25 July 2002

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/018250 A3

(54) Title: SINGLE NUCLEOTIDE POLYMORPHISMS IN GENES

(57) Abstract: The invention provides nucleic acid segments of the human genome, particularly nucleic acid segments from a gene, including polymorphic sites. Allele-specific primers and probes hybridizing to regions flanking or containing these sites are also provided. The nucleic acids, primers and probes are used in applications such as phenotype correlations, forensics, paternity testing, medicine and genetic analysis. A role for the thrombospondin gene(s) in vascular disease is also disclosed. Use of single nucleotide polymorphisms in the thrombospondin gene(s) for diagnosis, prediction of clinical course and treatment response, development of therapeutics and development of cell-culture-based and animal models for research and treatment are disclosed.

INTERNATIONAL SEARCH REPORT

Inte[illegible]al Application No
PCT/US 00/24503

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12Q1/68 C07K14/47 C07K14/78

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C07K C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
MEDLINE, SEQUENCE SEARCH, BIOSIS, EPO-Internal

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 5 750 502 A (KLAR AVIHU ET AL) 12 May 1998 (1998-05-12) SEQ ID NO:20 ---	1-30
A	POLYMEROPOULOS M H ET AL: "DINUCLEOTIDE REPEAT POLYMORPHISM AT THE HUMAN THROMBOSPONDIN GENE THBS1" NUCLEIC ACIDS RESEARCH, vol. 18, no. 24, 1990, page 7467 XP002188932 ISSN: 0305-1048 abstract --- -/--	1-30

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
5 February 2002	15. 05. 2002

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	van Klompenburg, W

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WANG D G ET AL: "Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome" SCIENCE, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE,, US, vol. 280, 1998, pages 1077-1082, XP002089398 ISSN: 0036-8075 the whole document ---	1-30
A	FAN J ET AL: "Genetic mapping: Finding and analyzing single-nucleotide polymorphisms with high-density DNA arrays" AMERICAN JOURNAL OF HUMAN GENETICS, UNIVERSITY OF CHICAGO PRESS, CHICAGO,, US, vol. 61, no. 4, SUPPL, 1 October 1997 (1997-10-01), page 1601 XP002089397 ISSN: 0002-9297 abstract -----	1-30

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 00/24503

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-30

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

Invention 1, claims 1-30

A method for predicting or diagnosing a vascular disease comprising; determining the nucleotide present at position 2210 of the thrombospondin 1 gene (SEQ ID NO:1). A nucleic acid molecule, a peptide (SEQ ID NO:2). A method for predicting or diagnosing a vascular disease comprising; determining the amino acid at position 700 of thrombospondin-1.

Invention 2, claims 31-60

A method for predicting or diagnosing a vascular disease comprising; determining the nucleotide present at position 2210 of the thrombospondin-4 gene (SEQ ID NO:3). A nucleic acid molecule, a peptide (SEQ ID NO:4). A method for predicting or diagnosing a vascular disease comprising; determining the amino acid at position 700 of thrombospondin-4.

Inventions 3 - 2547, claims 61-72

A nucleic acid molecule, an isolated gene product. A method of analyzing a nucleic acid sample. Every invention is characterised by each individual sequence of table 1 (corresponding to SEQ ID NO: 7-2551)

INTERNATIONAL SEARCH REPORT

Inte l Application No
PCT/US 00/24503

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 5750502 A	12-05-1998	US 5279966 A	18-01-1994
		AU 713198 B2	25-11-1999
		AU 1269897 A	15-05-1997
		AU 677185 B2	17-04-1997
		AU 3945593 A	08-11-1993
		CA 2133443 A1	14-10-1993
		EP 0670895 A1	13-09-1995
		JP 7508402 T	21-09-1995
		WO 9320196 A1	14-10-1993
		ZA 9302362 A	15-06-1994